

Errors in citation statistics

A curious absence from a list of 'hot papers' has led *Nature* to uncover some inaccuracies in the citation statistics compiled by the ISI. This adds to worries about relying heavily on these figures when rating scientific performance.

For some researchers, monitoring citation statistics and journal 'impact factors' is an intensely serious business. In many German universities, for instance, the impact factors of the journals in which scientists publish their work are tallied, and the data plugged into formulae that directly influence the funding given to individual departments. Worldwide, citation statistics are increasingly being used as a convenient metric to assess the quality of scientists' work.

Researchers with expertise in bibliometric analysis have long pointed out the potential pitfalls, and go to considerable lengths to validate the data used in their studies. But when citation statistics get into the hands of non-experts, rigour frequently flies out of the window. Comparing the impact factors of journals is meaningless, for instance, if they serve different disciplines in which widely different citation practices may prevail. But in the drive to rate scientists' performance, such comparisons are sometimes carried out.

The effective monopoly supplier of such statistics is the Philadelphia-based ISI, formerly known as the Institute for Scientific Information, and now owned by the Thomson Corporation. The ISI cannot be held responsible for the uses to which its statistics are put, and its web pages include a commentary on the limitations on citation data.

However, the ISI is accountable for the accuracy of the statistics it publishes. Subscribers to the ISI newsletter *Science Watch* will notice that its latest issue contains an apology to *Nature* and the International Human Genome Sequencing Consortium. In statistics published on the ISI's website, citation counts for the landmark paper describing the consortium's sequencing of the human genome (*Nature* **409**, 860–921; 2001) were so low that it was absent from the lists — of questionable use, even on a good day — of 'hot papers' in biology, which are published regularly in *Science Watch*.

When puzzled *Nature* staff examined the ISI's data, they found

that citations to the paper had been grossly undercounted. The same, it emerged, applied to several other prominent papers authored by a consortium, rather than by a list of individuals. To its credit, the ISI reacted promptly, amending its website and commissioning a journalist to write an account of the episode for *Science Watch*. The paper by the International Human Genome Sequencing Consortium now sits at the top of the hot papers list.

That would be that, were it not for further investigations that reveal other problems, this time with journals' impact factors — a measure of the average number of citations per paper. A journal's 2000 impact factor, for instance, is calculated by first counting the total number of citations made in that year, counted across all publications scanned by the ISI, to papers in that journal that were published in the preceding two years. This figure is then divided by the total number of items in the journal over the same two-year period that are deemed by the ISI to be citable — usually original research papers and review articles.

Preliminary studies by *Nature* have uncovered errors in the denominators used by the ISI that would seem to invalidate some previously published impact factors. For some of the journals published by the Nature Publishing Group, for instance, the number of citable papers tallied by the ISI has been inaccurate, leading to spurious variation in impact factor from year to year.

It is not the purpose of this article to put the record straight — that would require further detailed analysis. And for many scientists, the accuracy of a particular journal's impact factor will not be their most pressing concern. But these examples highlight an even greater need than previously realized — by us, at least, we confess — to check the ISI's data. Researchers, policy-makers and publishers who depend heavily on citation statistics should be urged to treat them with greater caution. And, it would seem, the ISI has some further investigation to do. ■

Lynch mob turns on lynx researchers

Biologists who tried to test the performance of a lab conducting genetic analysis have been unfairly pilloried.

Using hair from a captive lynx as a blind control to test laboratory genetic methods in a wildlife biology project seems like a reasonable procedure. But biologists in the US Pacific Northwest are now being subjected to vehement criticism for doing just that.

The seven biologists were part of a multi-agency team gathering hair samples in federal forests to try to determine the range of the threatened Canada lynx (*Lynx canadensis*). In three separate instances in 2000, different biologists sent in captive lynx hair to see if a US Forest Service laboratory in Missoula, Montana, could accurately identify samples from the species (see page 107). The biologists took this action, with the approval of their immediate superiors, after concerns were raised about the lab's work.

Now their attempts to test the rigour of the lynx project's genetic analysis have become fodder for interest groups and congressional

critics of US environmental-protection laws. The biologists stand accused of a "biofraud", and of conspiring with environmental groups to get vast tracts of forest placed off-limits for commercial and recreational use. In some erroneous media reports, the biologists are said to have "planted" the hair in the forest. Political rhetoric is flying like fur at a feline fracas.

This lynching is undeserved. The fact that the biologists felt compelled to take the action they did illustrates that the study's experimental design was deficient — blind controls should have been incorporated from the start.

Scientific leaders of the lead participating agencies, the US Fish and Wildlife Service and the Forest Service, must shore up their study methods. Clean data are needed to prevent years of court battles over use of the forests. The stakes — environmental, economic and recreational — are all too high for casual methodology to endure. ■





Frozen out
Argentina's turmoil
leaves scientists
fearing the worst
p104



Hot science
Researchers pitch in
to fight Australian
bush blazes
p105



Speedy verdict
Veterinarian under
fire for 'unnecessary'
feline drug study
p106



Arise Sir Jim
James Watson given
knighthood in New
Year Honours List
p108

Xenotransplant experts express caution over knockout piglets

Declan Butler

The safe transplantation of pig organs into human patients remains several steps from realization, experts say. Two announcements that research teams have cloned pigs lacking a gene involved in graft rejection still leave several obstacles to xenotransplantation intact, according to most specialists in the field.

On 2 January, PPL Therapeutics, the Scottish company that cloned Dolly the sheep, announced the birth on Christmas Day of five cloned knockout female piglets. Each had an inactivated gene for α -1,3-galactosyl transferase, an enzyme that adds the sugar α -1,3-galactosyl, or alpha-gal, to the surface of pig cells. The immune systems of humans and Old World primates, who lost this enzyme in evolution, recognize the sugar as foreign and kill transplanted pig organs in minutes.

The timing of PPL's statement was widely interpreted as an attempt to steal the thunder from a similar finding that was peer-reviewed and published by *Science* on 3 January (L. Lai *et al.* 10.1126/science.1068228) by groups at the University of Missouri at Columbia and



Porker potential: but the effects of fully deleting a gene that blocks xenotransplants are not known.

Massachusetts-based Immerge BioTherapeutics. Their four cloned piglets, all female, were born in September and October.

"The promise of xenotransplantation is

now a reality," Alan Colman, research director of PPL Therapeutics, claimed in his company's statement. But others find this contention premature. They point out that the elimination of the alpha-gal is incomplete in the new work — and that even if it were completely removed, it would by no means assure that xenotransplantation could succeed.

Each research team only knocked out one of the two copies of the gene — all of the piglets still make alpha-gal with the other copy. The companies intend to breed from their piglets and planned male litters to get offspring that have both genes knocked out and produce no alpha-gal. But the success of this step, which will take around 18 months, is not guaranteed. Jeffrey Platt, director of transplantation biology at the Mayo Clinic in Rochester, Minnesota, points out that although mice with double knockouts for this gene have been produced, this modification could prove lethal to pigs.

Nor is it clear that organs from pigs that lack alpha-gal altogether will be accepted by the human body. When pig organs are trans-



Alan Colman is confident of the pigs' prospects.

Clone pioneer calls for health tests

David Adam, London

Ian Wilmut, who led the team that created Dolly the sheep, has called for better assessments of the health of cloned animals, after revealing that Dolly has developed arthritis.

Wilmut, of the Roslin Institute near Edinburgh, says that Dolly's condition could have arisen because of genetic defects caused by the cloning process.

"What makes people slightly concerned is that Dolly has developed the arthritis in her hip and knee joint," Wilmut says. It is not uncommon for a five-year-old sheep to develop arthritis, but it usually strikes the elbows. Dolly is receiving anti-inflammatory drugs but may have to be destroyed if her condition deteriorates.

Wilmut says that other groups developing cloned animals could have failed

to reveal similar abnormalities. "I'm not sure that other companies are actually monitoring or publishing this information," he says. Wilmut suggests that independent assessment of cloned animal health and anonymous publication of the findings is needed before the technology is used on a large scale.

But other researchers insist that such information is already recorded and published. "I'm surprised, because I thought we had all been rather open," says Robert Lanza, a senior scientist at Advanced Cell Technology, based in Worcester, Massachusetts, a company that is also developing cloned animals. "We have just published a very comprehensive study where we analysed in extensive detail virtually all of our cloned animals fully from birth to adulthood," he says.

▶ planted into primates, a violent immune response is unleashed. This 'hyperacute rejection' destroys the blood vessels in the organ, cutting off oxygen, and turning it black within minutes.

Alpha-gal is known to play a major role in this acute response. But other foreign sugars will remain and the acute rejection may not be abrogated completely, says Robin Weiss, a virologist at University College London. "We won't know until we try," admits Ron James, PPL's managing director.

Should scientists succeed in overcoming hyperacute rejection, they will still face other rejection mechanisms including the 'delayed xenograft rejection' that occurs within days, when antibodies, macrophages and natural killer cells invade the organ. Data are insufficient to predict whether the absence of alpha-gal will alleviate these, says Fritz Bach, a xenotransplantation expert at the Harvard Medical School.

Alpha-gal also plays no role in the remaining major immunological barrier, T-cell-mediated chronic rejection, which can come into play months or years after the transplant. In human-to-human transplants this is controlled by lifelong immunosuppression, but this would probably be insufficient for xenotransplants which have a lot more potential targets for the immune system to attack. One promising approach, tricking the recipient into recognizing donor tissues as 'self', is being pursued by David Sachs, director of the Transplantation Biology Research Center at Massachusetts General Hospital.

Sachs is collaborating with Immerge BioTherapeutics on a system where cells from the thymus of the donor pig are first engrafted into the recipient while their immune system is temporarily disabled. As the system recovers, the recipient develops tolerance to the donor organs and tissues.

Even if all these hurdles are cleared, there is still the risk that animal viruses might jump to humans and cause man-made pandemics (see *Nature* **391**, 320–324; 1998).

Despite these problems James predicts that clinical trials of pig organs may be just four years away. Julia Greenstein, chief executive of Immerge BioTherapeutics, is more circumspect. "I'm saying that in three years time we want to have completed the preclinical trial in primates that would justify human trials," she says.

Financial analysts are also cautious — despite the jump in PPL's share price just after its statement was released. "There are huge difficulties and it will be a long way getting there," says Kevin Scotcher, a pharmaceuticals analyst at S G Cowen, the US-based investment bank. But Scotcher adds that a global market for transplant organs that could soon be worth \$6 billion annually will ensure intense investor interest in research related to xenotransplantation. ■



PABLO ANELLI/AP

Fiery protest: trouble on the streets of Buenos Aires is reflected in Argentina's laboratories.

Argentina's crisis heralds time of torment for scientists

Carol Marzuola

Argentinian scientists face a grim new year in the wake of their country's slide into political and economic chaos. With salaries unpaid and grant money effectively impounded by the banks, work in Argentina's labs is grinding to a halt. There are widespread fears that years of effort spent building globally competitive research groups could rapidly unravel.

"I think the economic situation can only worsen," says Armando Parodi, a biochemist at the National University of General San Martín's Institute for Biotechnology Research in Buenos Aires and a chief investigator for the national council for science and technology (CONICET). "The economic fiasco has brought on a social fiasco and the very cohesion of society has been affected."

Unfortunately for Parodi and other Argentinian scientists, it is unclear who in the government will shield them from the consequences of the crisis. As *Nature* went to press, the key post of national secretary of science and technology — which oversees CONICET, employing most of Argentina's leading researchers, and a parallel granting body, the national agency for the promotion of science and technology — remained unfilled.

The previous occupant, Adriana Puiggrós, resigned immediately after President Fernando de la Rúa, whose departure from office on 20 December brought the country's crisis to a head. Adolfo Rodríguez Sáa, the third of Argentina's five presidents since then, appointed a science and technology secretary during his one-week tenure, but the post was vacated when he resigned on 30 December. Science has not been a top priority for President Eduardo Duhalde, who took over on 2 January. "We're in a period of transition and uncertainty," says Juan Tirao, acting president of CONICET. Tirao suspects that the coun-

try's science agencies may be restructured in the wake of the economic and political crisis.

Argentina's scientists have struggled during the country's four years of recession. But their troubles deepened in early December when the government placed a 250-peso (US\$250) weekly limit on bank withdrawals, effectively denying access to research grants.

"We need the assistance of politicians and economists to avoid collapse and a waste of resources," says Osvaldo Sala, an ecologist at the University of Buenos Aires. Sala fears that the chaos will force many scientists to leave Argentina or to quit science altogether.

Monetary-exchange controls, imposed to protect the peso, have exacerbated the situation for scientists, many of whom rely on foreign equipment, reagents and journals. Sala, who heads a major project funded by the Inter-American Institute for Global Change Research, has been directly affected. He cannot transfer funds from an Argentinian bank to his colleagues in other countries.

Argentina's devaluation of the peso, announced on 7 January, pegs the official currency at 1.4 pesos to the US dollar, while allowing it to float on unofficial markets. But the devaluation itself is fraught with difficulty, and it is unclear, for example, whether imported scientific equipment could be bought at the official rate. "Things are changing by the minute," says Sala.

"A situation like this is so extreme that we can't predict what will happen in the next month," says biochemist Eduardo Olivero, director of the Southern Centre for Scientific Research in Tierra del Fuego. "Our hope is that the government will quickly clarify the situation for the country's scientific community — something that obviously hasn't been talked about in the past few days because of the gravity of our social and political crisis." ■

Legal move could open door to physics lab

Irwin Goodwin, Washington

High-energy physicists in the United States are hoping that a legal exemption could pave the way for construction of the proposed National Underground Science Laboratory.

Homestake, a disused gold-mine in Lead, South Dakota, owned by Canada's Barrick Gold, is the preferred site for the new laboratory.

But the company had said that without legal indemnity against future costs arising from environmental contamination at the site, it will flood the mine, eliminating its potential for use as a laboratory.

Just hours before Congress adjourned at the end of December, Senator Tom Daschle (Democrat, South Dakota) helped to arrange for the passage of a bill, the Homestake Mine Conveyance Act, that will remove Barrick Gold's legal liabilities and transfer the mine to the state of South Dakota. The legislation is likely to be signed into law shortly by President George W. Bush.

But officials at the National Science Foundation (NSF) — the agency most likely to pay for the laboratory — are unhappy that the exemption could land them with the bill for environmental problems at the site.

And the fact that it was won in large part through the efforts of Daschle, the majority leader in the Senate, could prove a mixed blessing for the project. Daschle is an early

front runner for the Democratic nomination for the 2004 presidential election, and his role in winning the exemption has come under attack from Republicans.

Several scientific advisory committees have identified the Homestake mine as the most promising site for an underground US laboratory, which would be used to study neutrinos and other phenomena of interest.

John Bahcall, an astrophysicist at the

Institute for Advanced Study in Princeton, New Jersey, and chair of one of these committees, notes that US researchers currently rely on overseas facilities for such work.

Bahcall says that researchers could "strike gold" at Homestake in such diverse fields as nucleon decay, solar neutrino measurements and the study of underground microorganisms. Reaching depths of 2,500 metres, Homestake is more than twice as deep as the overseas facilities, which means that it is even better protected from the cosmic rays that neutrino researchers seek to avoid by building their detection chambers underground.

The Department of Energy builds and administers most large physics facilities in the United States, but its current impoverishment has led researchers to look to the NSF as the most likely sponsor of the laboratory.

Robert Eisenstein, assistant director for mathematics and physical sciences at the NSF, agrees that Homestake is the best site for the laboratory. The NSF has reviewed the proposal twice and is considering spending \$280 million over five years on the project, although it has yet to make a final decision. It has been estimated that the lab might cost \$1.2 billion to build and operate for 10 years.

"The agency flatly refuses a stewardship role to protect the environment," says Eisenstein. "The project will need to take its place in a queue of proposals."



Homestake mine could house an underground lab where US physicists hope to strike gold.

Bushfires leave ecologists hot under the collar

Peter Pockley, Sydney

The bushfires that have raged across New South Wales are fuelling arguments over the Australian government's approach to fire research.

The fires have consumed around half a million hectares of bush and forest, prompting calls from ecologists for more work on measures such as controlled burn, which, they say, could help to limit the bushfire problem.

Currently, the government spends around A\$2 million (US\$1 million) each year on bushfire research — half of which goes to the Commonwealth Scientific and Industrial Research Organisation (CSIRO), and the rest to universities.

The small team at the CSIRO is already claiming some success in garnering information that may have saved lives during this summer's fires. And the team's leader, Jim Gould, says that a six-year study in Western Australia has yielded better information on the rate that fires spread under different wind conditions.

But Phil Cheney, who founded the



The burning issue: the devastation in New South Wales has led to calls for more fire research.

CSIRO's bushfire research team, has long campaigned for more fires to be lit under controlled conditions to reduce fuel levels in the bush. He points out that fire is natural in Australia, but that fuel now accumulates to dangerous levels so that the resultant fires "burn much more intensively".

Len Foster, chief executive of the Australasian Fire Authorities Council, says a concerted attempt to add to Australia's fire knowledge foundered last year, when a

consortium led by Melbourne and Monash universities failed to win funding for a Cooperative Research Centre dedicated to the topic. Gould says that the project failed because too few state fire authorities offered to match the federal government's support.

Science minister Peter McGauran concedes that Australia could "make a bigger effort than we are at present" in bushfire research, and promises to revisit the merits of the project.

Whale deaths caused by US Navy's sonar

Mark Schroppe

The US Navy has admitted that its use of a high-intensity sonar system caused a rash of whale strandings and deaths in March 2000.

Sixteen beaked and minke whales were found stranded on beaches in the Bahamas shortly after US Navy ships using the high-intensity sonar had passed by. Six are known to have died, and the rest were pushed back into the sea. But a fall in sightings of beaked whales has led researchers working in the area to believe that many more may have died. Autopsies on the animals revealed bleeding around the whales' inner ears and in one instance in the brain.

The Navy and the National Marine Fisheries Service (NMFS) responded to the incident by launching a series of investigations. An interim synopsis of the reports, released on 20 December 2001, concludes that the bleeding was caused by sound waves produced by the high-intensity sonar.

The synopsis was welcomed by environmental groups, which claim that high-intensity sonar may have caused other strandings. Autopsy evidence in previous



High and dry: the US Navy's use of high-intensity sonar left whales on the beach in the Bahamas.

cases could not identify the cause of the stranding. "This is the first time the Navy has really acknowledged responsibility for

anything like this," says Andrew Wetzler of the National Resources Defense Council, a New York-based environmental group.

But the report says that the high-intensity sonar may not pose a widespread threat to marine life. Such systems are commonly used, and the synopsis says the strandings were the result of unique local conditions. The sound waves were trapped in a layer of warm water, preventing their dissipation, and the whales could not escape because they were feeding in underwater canyons.

But according to Roger Gentry, coordinator of the NMFS acoustics team, the possibility that other conditions might cause similar problems cannot be ruled out, as it is not understood how the sound waves caused the bleeding. Several US research groups are looking into this issue.

Although he welcomes the report, Wetzler claims that the Navy is playing down the importance of its conclusions by focusing on unique characteristics of the event. "We really believe that this is an important piece of evidence that calls for all parties to re-examine all use of high-intensity sonar around the globe," he says.

The report comes at a sensitive time. The US Navy is seeking approval for a new high-intensity sonar system (see *Nature* 410, 505; 2001). The system, which operates at lower frequencies than that responsible for the Bahamas strandings, is needed to detect new, quieter submarines, the Navy says. The NMFS will have to approve use of the system under the Marine Mammal Protection Act, and Gentry says a decision should be made within a few months. But he says the report is unlikely to significantly affect the approval process. ■

♦ www.nmfs.noaa.gov

NIH faces action over HIV cat study

Erika Check, Washington

The US National Institutes of Health (NIH) is being sued over a research project that involves giving amphetamines to cats infected with the feline equivalent of HIV.

The Physicians Committee for Responsible Medicine (PCRM), a Washington-based pressure group that opposes animal experiments and advocates preventative health care, launched the action on 27 December. Using the Freedom of Information Act, it wants to force the NIH to release all the documents relating to the grant proposal made for the project.

The study is led by Michael Podell, a

veterinarian at Ohio State University in Columbus, who is investigating the interactions between drug abuse and AIDS.

Podell won a five-year grant worth \$1.7 million from the National Institute on Drug Abuse, part of the NIH, in autumn 2000 to study neurological changes in cats infected with feline immunodeficiency virus (FIV) and exposed to high doses of methamphetamine, the recreational drug more commonly known as speed.

The PCRM claims that the study is cruel and scientifically unnecessary, and wants the NIH to release more information about the project, including the outcome of Podell's pilot studies.

Podell, the NIH and Ohio State University have each declined to comment on the lawsuit. But in a previous statement, Podell said that his feline study will help to establish how HIV and drug abuse cause brain damage in human patients.

Murry Cohen, a psychiatrist affiliated with the PCRM, says that the feline model is too dissimilar from the human infection to yield clinically relevant information.

But Dennis Kolson, a neurologist at the University of Pennsylvania, says: "There are so few *in vivo* model systems for retroviral diseases like HIV that we have to utilize each one to its fullest extent, and the FIV system fits into that, as does the primate model." ■



Speed restrictions? Questions are being raised over a study that gives cats amphetamines.

MICHAEL PODELL

ANTONIO MIGNUCCI, CARIBBEAN MARINE MAMMAL LAB, UNIV. METROPOLITANA

Fur flies over lynx survey's suspect samples

Rex Dalton, San Diego

A study of the habitat of the threatened Canada lynx (*Lynx canadensis*) in US forests is embroiled in fierce controversy, after it emerged that wildlife biologists sent fur samples from captive lynx to a laboratory that was supposed to be monitoring the whereabouts of the animals in the wild.

The *Washington Times* published allegations last month that the biologists were seeking to distort a national survey of the lynx by planting the captive animals' fur in the forest.

Critics of wildlife-conservation measures in the United States — including powerful figures in both Congress and the Bush administration — have pounced on the allegations, claiming that they confirm their worst fears about how far government scientists will go to justify wildlife protection.

But the biologists involved hotly deny any intent to deceive, saying that the samples were never "planted" in the forest, but were sent to the laboratory to check that it was testing properly.

The existence of Canada lynx in the western United States is a highly charged political issue. Under the 1973 Endangered Species Act, areas that are populated by the threatened cat are subject to restrictions on logging, mining and public access.

The population has been monitored

since 1998 by a national lynx-detection programme, conducted by the US Department of Agriculture's Forest Service in cooperation with the US Fish and Wildlife Service and other federal and state agencies. Under the scheme, biologists place scratch pads in the forest to snag lynx hair, which is then submitted to a laboratory to confirm its identity.

In 1998, federal officials say, a private contractor hired to carry out the survey's first season found the lynx in several forest regions in Washington state. The next year, biologists at the Forest Service took over the study, but found the lynx in only one region of the state.

But in 2000, confusion arose among the various state and federal biologists who were sending samples for DNA analysis to the Forest Service's Forestry Sciences Laboratory in Missoula, Montana. In interviews and statements, several biologists questioned the laboratory's capabilities.

These concerns prompted seven wildlife biologists — in at least three independent instances at three separate agencies — to submit samples of captive lynx hair as 'blind controls'. "Everyone in the field was questioning the DNA analysis," says Jeffrey Bernatowicz, a biologist at the Washington State Department of Fish and Wildlife, "so people sent in control samples independently."

When these controls came to the attention of the Forestry Service over a year ago, the

agency hired a private Oregon firm to investigate. Last summer, the firm reported details of the control-sample methods, and the seven biologists were removed from the survey.

Although none of the 'control' specimens was included in the detection programme's data, the survey may now be discredited. The episode is being investigated by the inspector generals of two US government departments, and Congressional hearings are likely.



Lynx the loser: the plight of the endangered cat may be overshadowed by political jostling.

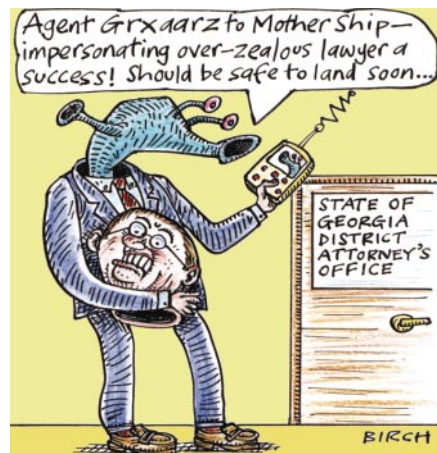
BBC WILD

Charges over computing project may set precedent

Erika Check, Washington

SETI@home users beware — aiding the quest for extraterrestrial intelligence could land you on the wrong side of the law.

Computer administrator David McOwen faces prosecution by the state of Georgia after he downloaded distributed-computing software — which divides time-consuming computing problems among many machines



— onto computers at DeKalb Technical College, where he worked two years ago.

Millions of users have installed the distributed SETI@home software, which analyses radio signals from space to search for signals from extraterrestrial life. McOwen had downloaded software from distributed.net, a website that specializes in software for tackling mathematical problems such as encryption.

McOwen, who is expected to face trial later this month, has been charged under Georgia's computer-hacking law. This law prohibits altering or interfering with computer data "with knowledge that such use is without authority". Similar laws exist in other states.

But David Joyner, McOwen's lawyer, says that his client did not violate any clear written policies. "It's inherent in his position to have the authority to run whatever program he felt like he needed to run," he says. Joyner adds that students at the college downloaded SETI@home software without being arrested.

The Electronic Frontier Foundation, a San

Francisco-based group that campaigns on civil-liberty issues related to information technology, says that the case — the first of its kind — has worrying implications. "We've had a number of people, primarily in the mathematical world, tell us this is scary, because there are clearly good things that can be done with distributed software," says Lee Tien, a lawyer for the organization.

One popular program, the Great Internet Mersenne Prime Search, lets users search for prime numbers of the form $2^n - 1$. A record-breaking four-million-digit prime number was found last November. Researchers at the University of Oxford are using a distributed computing project to screen small molecules for promising leukaemia drug candidates, and have signed up more than a million users since the project was launched last April.

Dave McNett, president and co-founder of distributed.net, denies that distributed computing projects are in danger. "I think this has raised awareness that you shouldn't run software on machines you don't own without permission," he says. "That has always been our policy."

Chinese ask German museum to return 'smuggled' fossil

San Diego The Chinese Society of Vertebrate Paleontology has asked a prominent German museum to return an important dinosaur fossil that may have been smuggled out of China.

Society officials in Beijing wrote to Fritz Steininger — director of the Naturmuseum Senckenberg in Frankfurt — on 25 December, seeking the repatriation of the psittacosaurid fossil, together with any other fossils that had been smuggled out of China. The Senckenberg purchased the specimen from a German dealer last summer for about US\$200,000 (see *Nature* **414**, 571; 2001). The fossil is unusual in that, unlike other specimens, its tail is clad with spines reminiscent of the quills on a porcupine.

The society called the trade in such illicit fossils "illegal and immoral".

Steininger says he has not yet read the e-mail from the society and declines to say if he would consider returning the fossil, adding that he will wait to discuss the issue with the society.

Watson's work hailed with an honorary knighthood

London US biologist James Watson, co-discoverer with Francis Crick of the double-helix structure of DNA, has received an honorary knighthood for "services to the UK–American partnership in the field of genetics and science".

Watson, now 73 and currently president of Cold Spring Harbor Laboratory in New York state, made the discovery with Crick nearly 50 years ago while at the University of Cambridge. The two published a now historic paper on their findings in 1953 (*Nature* **171**, 737–738; 1953) and went on to receive the Nobel Prize in Physiology or



Knighted: James Watson (left) has been honoured for services to UK–American science.

Medicine in 1962, together with Maurice Wilkins of King's College London.

Francis Crick is thought to have been offered a knighthood in the past but to have turned it down.

Gabriel Horn, a neuroscientist at the University of Cambridge who recently led an investigation into the causes of bovine spongiform encephalopathy, and Richard Brook, former head of the Engineering and Physical Sciences Research Council, a British funding agency, were also awarded knighthoods in the New Year Honours List.

Dawn and Kepler set for 2006 lift-off

Washington NASA has approved two new spacecraft for a 2006 launch, and allocated each a budget of US\$300 million.

The Dawn mission will orbit two large asteroids — Vesta and Ceres — reaching the first in 2010 and the second in 2014. Led by planetary scientist Christopher Russell of the University of California, Los Angeles, Dawn will be the first scientific mission to use ion propulsion, a technique pioneered by NASA's Deep Space 1 demonstration spacecraft. Kepler, a space telescope, will search for planets orbiting other stars.

Both spacecraft will be funded by NASA's Discovery programme, which emphasizes low-cost, highly focused Solar-System missions. NASA had been expected to fund one Discovery mission only. A third competing mission to study Jupiter's interior will not be funded (see *Nature* **409**, 124; 2001).

Kashmir dispute hits India's science links with Pakistan

New Delhi Tension over the disputed province of Kashmir is damaging the limited cooperation established between scientists in India and Pakistan.

Pakistani researchers due to travel next month to a conference on DNA fingerprinting in Hyderabad, India, have been refused visas by the Indian government. The conference has been organized by the Centre for Science and Technology of the Non-aligned and Other Developing Countries, based in New Delhi, which promotes scientific cooperation between its 40 member nations.

"India and Pakistan never signed a formal science-cooperation agreement but exchanged scientists on a case-by-case basis," says Y. P. Kumar, head of the international division at India's science ministry. "It is unlikely that this practice will be restored until political relations between the two countries have returned to normal."

Kumar adds that India is not currently allowing foreign researchers to conduct field studies close to the Pakistan border, which is an area of interest to seismologists.

Turkish stores battered by Californian earthquake

San Francisco Engineers at the University of California, Berkeley, spent the Christmas period completing a mock-up of Turkish grocery store, only to subject it to a simulated earthquake on 5 January.

The simulation was part of a project to investigate how wooden-framed buildings

respond to earthquakes. The three-storey building was built on a table capable of reproducing earthquake-like movements, and subjected to a recreation of the devastating quake that struck Izmit in Turkey in August 1999.

The project aims to reduce deaths and injuries associated with wooden-framed buildings by improving building standards.



Pretoria fights court demand for HIV drugs

Cape Town The South African government is to appeal against a High Court decision forcing it to provide HIV-positive mothers with a drug that helps to prevent mother-to-child transmission of the virus.

The court ruled on 14 December that all HIV-positive mothers had a constitutional right to free access to the antiretroviral drug nevirapine. The government is resisting supplying the drug, citing safety concerns and difficulties with distributing and administering it. Health minister Manto Tshabalala-Msimang was also ordered to present a national programme for preventing mother-to-child transmission to the court by 31 March.

The government says it will appeal on the grounds that the judgement interferes with its right to determine policy, and not because it does not support the provision of the drug. Nevirapine is currently being provided at 18 pilot sites throughout the country, but in only one province — the Western Cape — do the majority of HIV-positive pregnant women have access. The date of the appeal has not been set.

More editors quit over human cloning paper

London Two more editors of *e-biomed: The Journal of Regenerative Medicine* — the online journal which published a controversial paper on human cloning in November — have resigned from the editorial board, saying that the paper has no scientific merit. The paper claims to report the first cloned human embryo (J. B. Cibelli *et al.* *J. Regen. Med.* 2, 25–31; 2001).

Developmental geneticist Robin Lovell-Badge, of the National Institute for Medical

Research in London, and Davor Solter, director of the Max-Planck Institute for Immunobiology in Freiburg, resigned in December after another member of the editorial board, stem-cell pioneer John Gearhart, withdrew his name (see *Nature* 414, 570; 2001).

Lovell-Badge says that the data in the paper do not support the claims in it. "I think it was unfortunate that they published this without consulting us," he adds.

The journal's publisher, Mary Ann Liebert, has played down the resignations, saying that it is not uncommon for a member of an editorial board to resign.

Hawaiian laser puts a new 'star' in the sky

San Francisco A new 'star' was born on 23 December, thanks to the use of a Hawaii-based laser and the reflective properties of the Earth's atmosphere.

Astronomers at the island's Keck Observatory, together with researchers from the Lawrence Livermore National Laboratory in California, used a laser to illuminate a layer of sodium atoms some 95 kilometres above the Earth's surface. This created an orange glow, too faint for the human eye to see, around 1 metre in actual diameter.

The glow will be used as a 'guide star' to assist the observatory's adaptive optics technology. The system follows changes, caused by Earth atmosphere, in the brightness of a reference star. The telescope's optics are then adjusted to compensate for the changes. The new 'star' illuminates a spot in an area of the sky that has no natural reference star, and will allow many more real stars to be studied.

Keck researchers hope to have the laser fully integrated with the adaptive optics system by the middle of this year.

Shorter, brighter, better

A laser technology with military roots looks set to make a big impact on biology. By creating short pulses of intense radiation, free-electron lasers will advance our understanding of biological molecules. Navroz Patel reports.

It is a scene straight out of the movies — satellites, armed with high-powered lasers, blast incoming enemy missiles from the sky. But when President Ronald Reagan's controversial 'Star Wars' defence programme was phased out around a decade ago, this scene was relegated to the cutting-room floor.

Reagan's plans may be stalled, but the technology behind them is forging ahead. Free-electron lasers (FELs), which would have produced the intense beams needed to destroy missiles, are about to undertake a new mission — the transformation of structural biology.

In a few years' time, two new FELs could be producing X-ray laser pulses that are millions of times more intense than those from existing sources. Based at DESY, Germany's high-energy physics research centre in Hamburg, and the Stanford Linear Accelerator Center (SLAC) in Menlo Park, California, these devices have the potential to reveal the secrets of biological molecules that have so far resisted structural analysis.

The concept behind FELs dates back to the early 1970s, when it was proposed by John Madey, then a low-temperature physics PhD student at Stanford University in California (J. M. J. Madey *J. Appl. Phys.* **42**, 1906–1913; 1971). Many infrared FELs have been built in the intervening years, but until now interest in the technology has mainly been limited to the physics community.

FELs, like synchrotron X-ray sources, exploit the radiation emitted by fast-moving electrons as they change direction. Large FELs, such as those planned at DESY and SLAC, use particle accelerators to boost bunches of electrons to close to the speed of light. The bunches pass through an



This German particle accelerator forms the basis for a new generation of high-intensity lasers.

undulating magnetic field which forces the electrons to follow a sine-wave path.

This wiggling path causes the electrons to emit electromagnetic radiation, which quickly catches up, and interacts with, other electrons in the bunch. The effect of this interaction depends on the position of each electron and the phase of the wave that catches it up: some electrons are accelerated, whereas others are slowed down. The overall effect is to force the electrons into a series of smaller, more compressed bunches. The radiation emitted by these 'micro-bunches' is more intense, and helps to create more micro-bunches. The pulse that finally emerges from the FEL is a series of high intensity bursts from several thousand micro-bunches given off as they leave the magnetic field.

Racing pulses

One of the biggest selling points of the new FELs is the intensity of radiation this process produces. "The X-ray FELs will offer 10 orders of magnitude more photons per pulse than the most intense synchrotrons currently available," says Janos Hajdu, a biochemist at Uppsala University in Sweden, who has worked on the plans for both new FELs.

Operations at DESY will halt in the next few months to allow for the upgrade, which should be finished in 2004. Keith Hodgson, director of synchrotron research at SLAC, says Stanford's FEL should be ready in late 2006.

The two new devices will initially produce X-ray pulses with wavelengths of around 10 nanometres but should eventually operate in the 0.1-nanometre range. Last September, researchers at DESY took a step towards this goal when they coaxed ultraviolet radiation out of their existing device. At 80 nanometres, the light's wavelength was the shortest ever produced by a FEL.

Both facilities will appeal to researchers from many disciplines. Physicists are keen to use the new FELs, as the beams will let them study little-understood states of matter. Warm dense matter (WDM), a state somewhere between a solid and a plasma, is one example. WDM is found in the cores of large planets, but can be created in the laboratory by focusing the high-intensity radiation produced by FELs on metal targets. The character of the WDM can then be probed by observing the way it scatters laser light.

But it is in biology that the X-ray FELs will have the biggest impact. Unlike radiation of

longer wavelengths, X-ray pulses are diffracted by the gaps between atoms in molecules. By studying the patterns made by these diffracted rays, biologists can deduce the structure of the molecule under analysis.

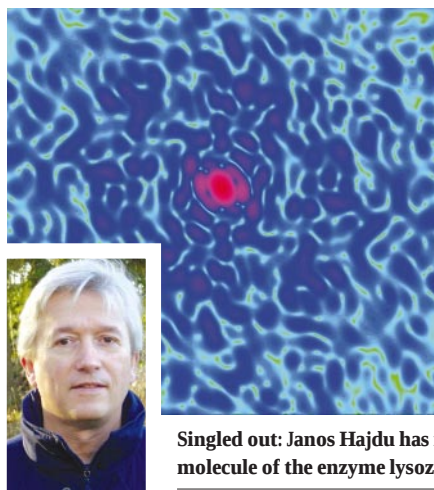
X-rays produced by conventional lasers and synchrotrons are already used to probe the structure of biological molecules, but both systems are limited by the intensity and pulse length of their radiation. The short, high-intensity pulses from the new FELs should allow researchers to overcome both of these stumbling blocks.

X-ray diffraction works well with biological molecules in crystalline form, as the rays are reflected between the crystal's layers and combine to produce a measurable diffraction pattern. This means relatively low-intensity X-ray sources can be used to create diffraction patterns. But for single molecules, low-intensity X-rays are of little use, as the signal produced is too weak and noisy to interpret.

The new FELs will allow analysis of samples of just tens of molecules, and perhaps even single molecules, which should remove a host of limitations. The shapes of proteins in a crystal, for example, are constrained by the surrounding molecules, so current diffraction studies can probe only a limited range of the shapes that the molecules are thought to assume. And because the shapes of proteins in a crystal vary, X-ray diffraction actually records an average of a range of different structures. It is also often impossible to study a protein's interactions with other biological molecules when it is frozen in crystalline form.

At the moment, structural biologists use nuclear magnetic resonance spectroscopy to overcome these obstacles. This can reveal the structures of proteins in solution, but it has not yet been made to work for large proteins. Hajdu believes this obstacle will be removed once the X-ray FELs are up and running. "The new FELs could get rid of the 'crystal' in 'crystallography' altogether," he says. In addition, the frequency of FEL radiation can be adjusted, allowing researchers to tailor the pulse to suit the molecule being studied.

Hajdu has already used computer simulations to reveal what the X-ray pattern created



Singled out: Janos Hajdu has modelled X-ray diffraction of an individual molecule of the enzyme lysozyme (left) and a virus that affects tomatoes (right).

by both a single enzyme and an individual virus will look like (R. Neutze, R. Wouts, D. van der Spoel, E. Weckert & J. Hajdu *Nature* **406**, 752–757; 2000). Once available, he intends to use the new FELs to test these predictions out.

Beamline bonanza

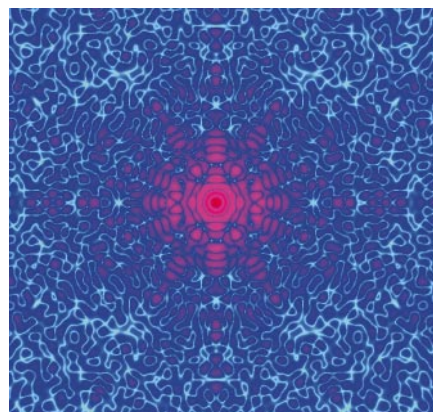
Other researchers plan to use the FELs to study the proteins that crystal-based methods have missed altogether. Proteins linked to lipids and embedded in cell membranes, for instance, are very difficult to crystallize. At present, says Carol Robinson, a chemist at the University of Cambridge, there is relatively little good structural data on membrane proteins. These proteins are important, she says, because they regulate traffic into and out of cells.

Another advantage offered by FELs is their short pulse length. Synchrotrons emit pulses that are typically a few tens of picoseconds (10^{-12} seconds) long. But vibrating molecules can change shape in a few femtoseconds (10^{-15} seconds). The femtosecond pulses produced by FELs should open up these vibrations for study. Indeed, among the other FELs under development, one will soon be using short infrared pulses to study the vibrations of DNA.

Researchers at the University of Maryland in College Park are building a 'compact' FEL — a smaller device that uses lower-speed electrons and so does not need a large particle accelerator. Compact FELs produce pulses that have longer wavelengths and a lower intensity than those from X-ray FELs.

The Maryland FEL, which should be ready this year, will be used to study how changes in the shape of DNA affect the way it interacts with other molecules. The vibrations within DNA and the larger changes in the shape of the molecule itself are thought to play a fundamental role in the synthesis of other biological molecules. A clear-cut experimental analysis of this role has so far eluded researchers — but the Maryland FEL should help to redress the balance.

With a wavelength of between 3 and 30 micrometres, the FEL's radiation is the right



E. WECKERT/DESY



Clear cut: the high-intensity beams from free-electron lasers have been used in eye operations.

frequency to make DNA vibrate by interacting with the concentrations of charge found along DNA's double helix. Researchers can then follow interactions between the vibrating DNA and other molecules by studying the way additional FEL pulses are scattered.

Compact FELs also have medical uses. Eye surgeons at Vanderbilt University in Nashville have used short, intense FEL pulses to operate on the back of the eye.

Given that FELs offer a range of advantages over other light sources, why are they not more widely used? Conventional lasers, by contrast, are found everywhere from industrial manufacturing to domestic compact disc players.

Madey, now at the University of Hawaii in Manoa, suggests the problem is that FELs have largely been the preserve of accelerator physicists, who are less in tune with potential applications. "They have worked without the scientific insights needed to adapt the technology for use in near-term research and manufacturing applications," he says.

But as excitement about the new facilities grows, the profile of FELs is set to get a much-needed boost. Star Wars, it seems, may have a surprisingly useful legacy.

Navroz Patel is a freelance writer in New York.

DESY ♦ tesla.desy.de

SLAC ♦ www-ssrl.slac.stanford.edu/lcls

D. JOHNSON/VANDERBILT UNIV.



Keith Hodgson hopes to have Stanford's X-ray free-electron laser up and running by 2006.

Betting on tomorrow's chips

At the proteomics frontier, dozens of companies are trying to develop the protein equivalent of DNA microarrays. But designing these chips poses much tougher technical challenges, says Alison Abbott.

The genomics revolution has already brought us DNA chips — microarrays that consist of short DNA sequences immobilized on a surface. By determining which spots bind to messenger RNA (mRNA) extracted from a biological sample, geneticists can obtain an instant snapshot of the activity of thousands of genes at a time.

DNA microarrays are transforming studies of gene expression. But some scientists are already dreaming of the chips of the future, which they argue will carry tens of thousands of protein 'capture' molecules, each geared to identify and bind to one particular protein. With an appropriate detection system, such chips would be even more valuable than their DNA counterparts. Proteins, after all, are the business end of gene expression and the usual target of drugs. A chip that could simultaneously analyse the production of tens of thousands of them would be a boon, both to those engaged in fundamental research and to the drug industry.

Dozens of companies are already working on technologies to make the chips, and the hype is feverish. But sober analysis indicates that the technical hurdles to be overcome are so great that over the next couple of years we can only expect chips with fewer than 100 capture molecules; the handful of chips marketed so far carry fewer than 10.

Such low-density chips are fine for certain

applications, such as simple medical diagnostics. But for large-scale proteomics projects that aim to determine how complex patterns of protein production vary with disease, they are inadequate. And although some companies claim to be making good progress towards chips with tens of thousands of capture molecules, many experts are sceptical.

"The hype from extrapolation from DNA arrays is very harmful — the expectations are too high," says Richard Mason, a business-alliances analyst with the British company Cambridge Antibody Technology (CAT). "Proteins are much more complicated, and development costs will be orders of magnitude greater," he argues.

The protein poser

For DNA chips, designing capture molecules and developing read-out systems were relatively easy. Strands of DNA bind tightly and specifically to mRNAs with a 'complementary' sequence. And if the mRNAs in a sample have all been tagged with a fluorescent dye, determining where on the array they have bound is also simple.

Proteins pose a much tougher challenge. Specific capture molecules must be designed for all possible proteins encoded by the genome — and also for the modified forms produced by processes such as phosphorylation or the addition of sugar groups.

Whereas the binding between DNA and mRNA is highly specific, finding a capture molecule that will bind with high affinity to one protein alone is extremely difficult.

The classical capture molecules for proteins are antibodies, which are themselves proteins, and most of the companies in the protein-chip business are working with them. "So far, antibodies are the only capture molecules that have been demonstrated to work at high specificity and sensitivity," says Leigh Anderson, chief scientific officer of Large Scale Biology, which has its proteomics division in Germantown, Maryland.

Nowadays, large libraries of antibodies can be produced using a procedure called phage display. CAT, which has been manufacturing antibodies since 1990, has cloned billions of distinct antibody genes from white blood cells of healthy individuals, and has inserted them into viruses called phages that infect *Escherichia coli* bacteria¹.

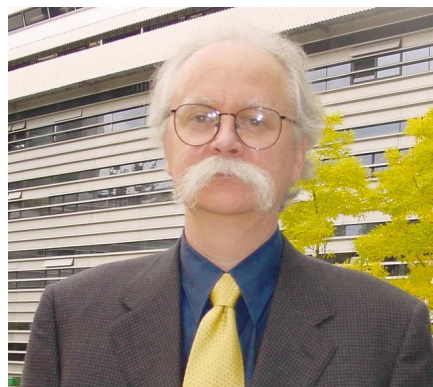
The phages reproduce in cultures of the bacteria. Infected cells eventually rupture, releasing phages into the growth medium. The remaining bacterial cells can be centrifuged away, giving a soup of phages, each of which carries on its surface the antibody encoded by the inserted gene. CAT has developed libraries that contain ten billion phage antibodies, says Kevin Johnson, the company's chief technology officer. Antibodies that bind to a particular target are then fished, or 'panned', from this soup, typically using a plastic surface on which the protein in question has been immobilized.

Capture molecules on a protein chip need to bind with high affinity because some of the most interesting proteins in a biological sample — such as hormones, growth factors and intracellular signalling proteins — are present only at very low concentrations. In practice, capture molecules must be able to identify target proteins from nanomolar (10^{-9} molar) to picomolar (10^{-12} molar) solutions.

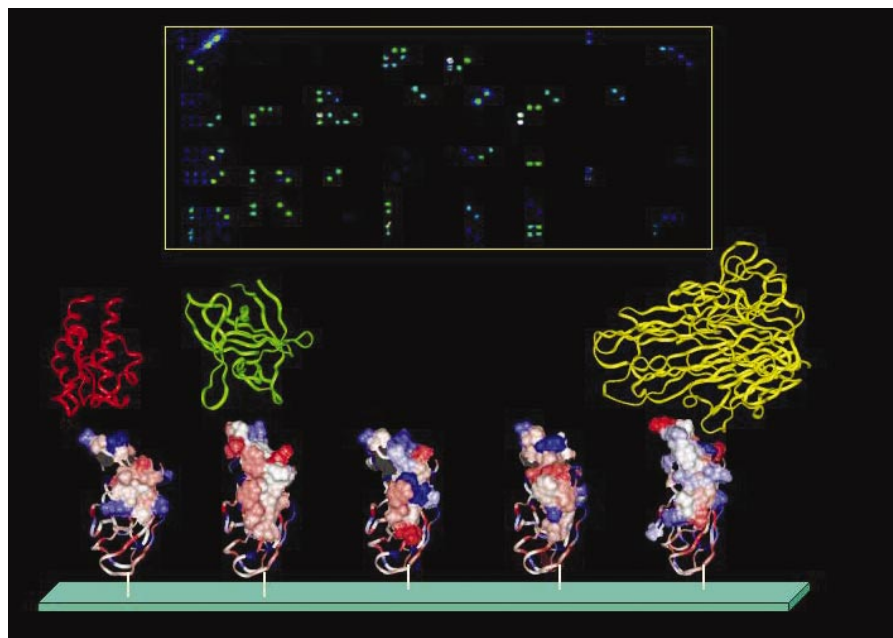
Success in identifying such molecules for large numbers of proteins is simply a matter of statistics — the bigger the library, the greater the chance of finding high-affinity antibodies. CAT has been successful in identi-



Turn up the volume: Kevin Johnson says CAT has created libraries of some ten billion antibodies.



Leigh Anderson of Large Scale Biology backs antibodies as the most reliable 'bait' molecules.



The protein-chip concept: capture molecules immobilized on a surface (bottom) bind to specific target proteins (here in red, green, yellow), giving a signal that can be read by viewing the chip (top).

fying specific antibodies that bind with high affinity to individual targets — for example, it has an antibody against tumour-necrosis factor- α (ref. 2) that is now in advanced clinical trials as a treatment for rheumatoid arthritis. But routinely coming up with high-affinity antibodies for thousands of target proteins is a different matter. Bigger libraries are needed, but the cumbersome step of fermenting bacterial cultures in phage display makes it difficult to expand libraries further.

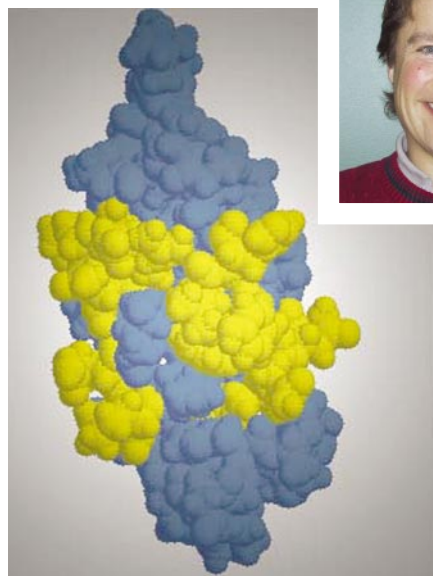
So CAT is now turning to a new 'ribosome display' technique developed by the Delaware-based company Aptein, which it purchased in 1998. Here, the phage is substituted with a ribosome, the cellular machinery that translates mRNA into protein. Normally, an mRNA molecule passes through the ribosome like ticker-tape and is released, along with the newly synthesized protein molecule, when a sequence of three bases known as a 'stop codon' is reached. In Aptein's technology, stop codons are eliminated so that the completed antibody and its mRNA remain bound together on the ribosome. The system, which CAT is now optimizing, is entirely cell-free and so is more amenable to automation. This should make it possible to construct libraries that are orders of magnitude larger than those created using phage display.

But antibodies have their drawbacks. They tend to be denatured — lose their structure — when heated or exposed to other stresses such as changes in pH. And companies that are trying to enter the field are also restricted by that fact that many key antibodies and antibody technologies are covered by patents. So newcomers have been driven to look for alternatives.

One leading contender is 'combinatorial

engineering' of protein scaffolds. A scaffold is a domain of a large protein that, like an antibody, can be made to bind to an enormous range of other proteins by subtly changing its sequence of amino acids. The scaffolds are more stable than antibodies to heat or other stresses, and are sufficiently small that they, or the genetic sequences to encode them, can be synthesized chemically.

A handful of competing systems are being developed by various start-up firms. But two companies — Affibody and Phylos — are currently battling for supremacy in this area. Each claims that its proprietary scaffolds will allow the development of chips that can analyse tens of thousands of proteins simultaneously.



Stefan Ståhl (inset) says that 'affibody' scaffold domains can bind to a wide variety of proteins.

Phylos, based in Lexington, Maryland, uses a 100-amino-acid domain from a structural protein called human fibronectin for its scaffold. Like the binding sites of antibodies, this domain consists of a rigid unit that supports loops of varying lengths. Small changes in amino-acid sequence change the shape of these loops, and hence the structures to which the domain will bind. By substituting different amino acids in the loops held by a scaffold, trillions or more variants can be made. Libraries of these can then be searched to identify any that bind to a particular protein.

Phylos has created vast libraries of scaffold variants, which it calls trinectins, using a proprietary mRNA-display system. This is a cell-free system similar to CAT's ribosome display, except that the mRNA remains bound to the proteins produced and the ribosomes are washed away³. "It is a very simple process and easy to automate," says Richard Wagner, Phylos's head of R&D.

Using this system, Phylos produces libraries of some 10^{13} variants. By applying several rounds of panning and washing away loosely bound proteins at each round, Wagner claims to have identified molecules that can capture their targets with nanomolecular or higher affinity — although nothing has yet been published. Targeted proteins include immune signalling molecules and their receptors.

Winning combinations

Affibody, based in Stockholm, Sweden, uses as its scaffold a domain of staphylococcus protein A, a bacterial surface protein that consists of 58 amino acids⁴. It normally interacts, through a binding surface made up of 13 of its amino acids, with immunoglobulin G molecules — the main class of antibodies in the blood. But by substituting these 13 amino acids, either singly or in combination, 'affibody' scaffolds can be made to bind to a wide variety of other proteins — in theory, it should be possible to create 10^{16} different affibodies. And since 1998, using phage display, Affibody has developed libraries containing up to 10^8 variants.

"We have never failed to find a binder for the proteins we throw into our soup of variants," says Stefan Ståhl, Affibody's chief scientific officer. Again, the company conducts repeated rounds of panning, and has developed affibodies with nanomolar affinities for several proteins⁵.

Given the limitations of phage display, Affibody is looking for a cell-free alternative to boost library sizes. "We are working on three different selection systems which do not involve passage of phages through a bacterial system and which are more suited to automation," says Ståhl. He declines to give

details, citing commercial confidentiality.

But not everyone is convinced that approaches that rely on *in vitro* screening for high-affinity binders will deliver the goods. Apart from the difficulty of generating sufficiently large libraries, another problem is that the selected molecules might cross-react with other proteins. "I think the body is a better vehicle for screening than a library," says Ian Humphery-Smith of the University of Utrecht in the Netherlands, who has founded a company called Glaucus Proteomics.

Humphery-Smith is instead deriving antibodies using mice engineered to have a human immune system. Each mouse is immunized with multiple antigen proteins, and after 40 days its blood plasma is screened against a protein chip with the same antigens attached. "In this way we can screen for cross-reactivity of antibodies as well as



Richard Wagner says Phyllos's system is easy to automate.

specificity and high affinity, in one go," says Humphery-Smith. The system will spawn a chip containing 150,000 antibodies within two-and-a-half years, he claims.

Other companies have eschewed proteins as capture molecules, and have turned instead to 'aptamers' — short strings of DNA or RNA that constitute specific binding partners for a range of proteins. Work on aptamers is at an earlier stage than that with protein-based capture molecules, but if the technology can be made to work, it has strong appeal — aptamers can be synthesized chemically and the chips could be produced using the same high-throughput techniques already perfected for DNA microarrays.

"It's a fascinating idea," says Anderson of Large Scale Biology. But he warns that aptamers have not yet been shown to bind specifically to individual proteins when confronted with a complex mixture.

The leader in the aptamer field is SomaLogic of Boulder, Colorado, which is collaborating with a major player in proteomics, Celera Genomics of Rockville, Maryland. SomaLogic has produced a library of 10^{15} DNA molecules in which thymidine — one of the four 'letters' of the genetic code — is replaced with bromodioxuracil. Already, the company has identified aptamers that bind with high affinity to a range of target proteins. "We expect to have a thousand proteins on a chip by next summer," says Larry Gold, SomaLogic's chief executive officer.

In SomaLogic's 'photoaptamer' system⁶, one end of each aptamer is covalently bound to the chip surface. Target proteins in a sample are captured by their individual aptamers, and the chip is then exposed to ultraviolet light, which causes the bromodi-

Some companies are starting to think about who their customers will be, and what they will expect.

oxuracil to cross-link with the captured proteins. Unbound proteins can then be washed away and the remainder can be identified using a general protein stain.

This simple detection system would solve the second major technical challenge of protein chips — creating a reliable and rapid read-out system. General protein stains cannot be used in systems in which the capture molecules are also proteins. And, unfortunately,



Larry Gold's SomaLogic leads the aptamer field.

simply labelling all the proteins in a sample with a fluorescent tag, as with the mRNAs detected by conventional DNA microarrays, is not a viable option, because different proteins take up the tags to different extents.

Currently, most systems rely on adding labelled antibodies to the proteins after they have been captured on the chips. But these 'sandwich assays' are hard to use as it is difficult to get sufficiently large numbers of antibodies into solution, and antibodies may bind to non-target proteins.

Sandwich rapped

For this reason, companies working with protein-based capture molecules are trying to develop systems that do not involve sandwich assays. One idea is to label the capture molecules so that they will signal when they have bound to their target protein. Affibody, for instance, is working on a system dubbed FLAME, which is based on a technique called fluorescence resonance energy transfer. In this system, the affibody capture molecules are engineered to include two tags, one of which fluoresces but is suppressed by the second if it is close by. When an affibody binds to its target, the two tags move away from one another and fluorescence occurs⁷.

Other approaches rely on detecting the physicochemical changes that occur when a capture molecule binds to its target molecule. Several companies, including BiaCore of Uppsala, Sweden, are developing a method called surface plasmon resonance, which detects differences in refractive index at the surface of a capture molecule⁸. The Dutch electronics giant Philips, based in Eindhoven, is collaborating with Humphery-Smith to measure the changes in potential difference between the chip surface and the sample solu-

tion that accompany binding. It is also using microelectromechanical devices to record the physical changes associated with binding.

SomaLogic, meanwhile, says that it is refining its own simple detection method to cope with another problem — the fact that the concentrations of different proteins in a biological sample can vary over several orders of magnitude. Detecting them all may require samples to be split, diluted to different extents, and analysed repeatedly on duplicate chips — or, alternatively, splitting capture molecules for high- and low-concentration proteins between different chips. SomaLogic believes that it can solve the problem using single chips, but so far it has revealed no details.

As companies work to resolve the technical problems of making protein chips viable, some are also starting to think about who their customers will be, and what they will expect.



Cell biologist Gavin MacBeath will use the chips.

"There really has been too little concentration on what the customers want, and too little consideration of what they will be prepared to pay," says Johnson of CAT.

Unless there is a significant technical advance, argues Johnson, the cost of protein chips will increase geometrically with the number of proteins on each chip. This is why CAT has, for now, opted for low-density chips, each addressing custom-made subsets of proteins.

Such chips will find their uses — in basic biology as well as diagnostics. "A chip that could follow a few dozen proteins in a cell-signalling pathway would be a great way to unravel the whole pathway," says Gavin MacBeath of Harvard University's Bauer Center for Genomics Research. "We could see which proteins upregulate in compensation when a particular target protein, like a receptor, is blocked."

As for which — if any — technology can provide the genuine proteomics breakthrough that would give rise to chips carrying tens of thousands of protein-capture molecules, investors are now placing their bets. ■

Alison Abbott is Nature's senior European correspondent.

1. Vaughan, T. J. *et al. Nature Biotechnol.* **14**, 309–314 (1996).
2. Mahler, S. M., Marquis, C. P., Brown, G., Roberts, A. & Hoogenboom, H. R. *Immunotechnology* **3**, 31–43 (1997).
3. Roberts, R. W. *Curr. Opin. Chem. Biol.* **3**, 268–273 (1999).
4. Nord, K. *et al. Nature Biotechnol.* **15**, 772–777 (1997).
5. Nord, K. *et al. Eur. J. Biochem.* **268**, 4269–4277 (2001).
6. Brody, E. & Gold, L. *Rev. Mol. Biotechnol.* **74**, 5–13 (2000).
7. Karlstöm, A. & Nygren, P.-A. *Anal. Biochem.* **295**, 22–30 (2001).
8. Nelson, R. W., Nedelkov, D. & Tubbs, K. A. *Electrophoresis* **21**, 1155–1163 (2000).

CAT ♦ www.cambridgeantibody.com

Phyllos ♦ www.phyllos.com

Affibody ♦ www.affibody.com

Glaucus Proteomics ♦ www.glaucusprot.com

SomaLogic ♦ www.somallogic.com

BiaCore ♦ www.biacore.com

Theoretical models of sheep BSE reveal possibilities

But we must remember that these theories are based on speculation, not on fact.

Sir—We do not yet know whether bovine spongiform encephalopathy (BSE) entered the UK sheep population at the height of the epidemic among cattle and, if it did, what the likely effects would be or whether the BSE prion could have persisted in sheep. The one study aimed at illuminating this question has collapsed in what might generously be called a failure of experimental protocol (see *Nature* **413**, 760; 2001). But if BSE did get into the sheep flock, it seems reasonable to assume that cases of variant Creutzfeldt–Jakob disease (vCJD) could result among humans.

Two new theoretical papers, one published in *Science* and the other in *Nature*, describe a range of possibilities for what might have happened in the past and what might be going on today.

Kao *et al.* (*Science* 23 November 2001; 10.1126/science.1067475) use such data as are available on the amount of meat and bone-meal sheep ate and the dose–response relationship. They estimate that about 80% of the UK flock is susceptible to BSE, with variations among genotypes and locations. They conclude that between 10 and 1,500 sheep may have become infected in 1990 and that with limited maternal transmission fewer than 20 clinical cases a year of BSE (hidden among 10,000 or so sheep with scrapie) are now occurring. They suggest that a much larger epidemic in sheep could develop in the coming decades if BSE transmits horizontally, but their paper does not investigate this possibility in detail.

Ferguson *et al.* (*Nature* 9 January 2002; 10.1038/nature709), in contrast, focus on horizontal transmission and the development of the epidemic in the next 80 years. They estimate the 95% confidence interval for future vCJD mortality to be 50–50,000 deaths considering exposure to bovine BSE alone, with the upper bound increasing to 150,000 once exposure to the worst-case ovine BSE scenario examined is included. Yet this statement phrased in terms of confidence intervals may unintentionally convey to readers an unjustified degree of statistical precision—an issue discussed by Graham Medley (*Science* **294**, 1663–1664; 2001). These estimates derive from a valuable but speculative exploration of nonlinear epidemiological models whose parameters are no more than illustrative guesses. Some simplifying assumptions in the model of Ferguson *et al.*, for example, are that infected flocks become immune for 20 years and that genetic variation in resistance is among rather than within flocks (one-third are assumed susceptible). These assumptions could have substantial

implications for projections of what could happen.

The crucial question for policy-makers is what should be done to manage a purely theoretical risk that BSE is present in sheep? The draft UK government contingency plan, should BSE be found in the present-day flock, can be seen at www.defra.gov.uk/corporate/consult/sheepbse/index.htm. Throughout Europe, current practice is to remove from the food chain sheep tissue that would carry infectivity if it was present (such as the brain and spinal cord of older sheep). But studies of experimental infection of sheep with BSE show that the current strategy still allows tissues that could harbour BSE to enter the food chain. Ferguson *et al.* consider the likely effects of removing more tissues and of earlier-age cut-offs. Again, given the uncertainties and assumptions, the quantitative details cannot be taken literally, although the

theoretical risk might be reduced substantially by allowing only lambs under six months old into the food chain and by removing more of the tissues that could harbour infectious material. The option of testing sheep for transmissible spongiform encephalopathies as they enter the abattoir and keeping infected animals out of the food chain is still some way off, as a cheap test of proven reliability does not yet exist.

In the light of such scientific uncertainty there is no easy formula for deciding on the right level of precautionary risk management. Studies such as those of Kao *et al.* and Ferguson *et al.*, which make varied theoretical assumptions, are useful in an open debate about policy. But their tentative nature must be kept in mind and the many uncertainties acknowledged.

**John R. Krebs, Robert M. May,
Michael P. H. Stumpf**

*Department of Zoology, South Parks Road, Oxford
OX1 3PS, UK*

Dropped genetics paper lacked scientific merit

Sir—Even though the controversial withdrawal of a paper on the genetic relatedness of Palestinians and Jews by the journal *Human Immunology* (see *Nature* **414**, 382; 2001) is a minor episode compared with the tragedies caused by ethnic/religious conflicts over past decades, the issues involved are worth revisiting.

The stated purpose of the paper by Antonio Arnaiz-Villena *et al.* was to “examine the genetic relationships between the Palestinians and their neighbours (particularly the Jews) in order to: (1) discover the Palestinian origins, and (2) explain the historic basis of the present ... conflict between Palestinians and other Muslim countries with Israelite Jews”.

They conclude: “Jews and Palestinians share a very similar HLA genetic pool that supports a common ancient Canaanite origin. Therefore, the origin of the long-lasting Jewish–Palestinian hostility is the fight for land in ancient times.”

It is difficult to believe that knowledge of genes may help to explain the present conflict. Although population genetics can address issues of relatedness of populations, mating patterns, migrations and so on, obviously it cannot provide evidence about reasons for conflicts between people.

Our primary concern, however, is that the authors might be perceived to have been discriminated against for political, as

opposed to legitimate scientific, reasons.

Even a cursory look at the paper's diagrams and trees immediately indicates that the authors make some extraordinary claims. They used a single genetic marker, HLA DRB1, for their analysis to construct a genealogical tree and map of 28 populations from Europe, the Middle East, Africa and Japan. Using results from the analysis of a single marker, particularly one likely to have undergone selection, for the purpose of reconstructing genealogies is unreliable and unacceptable practice in population genetics.

The limitations are made evident by the authors' extraordinary observations that Greeks are very similar to Ethiopians and east Africans but very distant from other south Europeans; and that the Japanese are nearly identical to west and south Africans. It is surprising that the authors were not puzzled by these anomalous results, which contradict history, geography, anthropology and all prior population-genetic studies of these groups. Surely the ordinary process of refereeing would have saved the field from this dispute.

We believe that the paper should have been refused for publication on the simple grounds that it lacked scientific merit.

**Neil Risch*, Alberto Piazza†,
L. Luca Cavalli-Sforza***

**Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA*

†Department of Genetics, Biology and Biochemistry, University of Torino, Via Santena 19, 10126 Torino, Italy

Why natural may not equal healthy

Many believe that the natural toxins in their food are safer than synthetic ones.

Naturally Dangerous: Surprising Facts about Food, Health, and the Environment

by James P. Collman

University Science Books: 2001. 280 pp. \$29, £19.99

John Krebs

If you ask people what they fear about their food, typically the top half-dozen concerns are food poisoning, BSE, growth hormones used in animals, animal feed, pesticides and genetically modified (GM) food (www.food.gov.uk). But how do these perceived risks stack up with the estimates of deaths caused by food? Acknowledging that these are only approximate, and that great uncertainties surround some of the numbers, two food risks tower above the rest — the dietary contributions to cardiovascular disease and to cancer. These risks, taking a fairly conservative estimate, probably account for more than 100,000 deaths per year in Britain. Food poisoning probably accounts for between 50 and 300 (similar in range of magnitude to the risk of choking to death on food or suffering a fatal accident while getting into or out of bed). As far as we know, growth hormones (banned in Europe) and pesticides in food, as well as GM food, are not responsible for any deaths.

A generally accepted psychological explanation for the discrepancy between perceived and actual risk is the one based on Paul Slovic's identification of the range of factors that make risks seem more frightening. Thus, for example, risks that are under someone else's control, potentially catastrophic and unfamiliar are perceived as greater than those with the opposite features. That is why most of us view riding our bicycle in a busy street as a more acceptable risk than living near a nuclear power station, although rational analysis says that you should stay off your bike.

James Collman writes about another important dimension of risk perception — naturalness: "Many Americans are under the mistaken impression that if something is 'natural' it is safe." As far as food is concerned, Collman covers similar ground to that in Julian Morris and Roger Bate's *Fearing Food* (Butterworth-Heinemann, 1999) and Douglas Powell and William Leiss's *Mad Cows and Mother's Milk* (McGill-Queens University Press, 1997).

Perhaps one of the most telling arguments against the 'natural equals safe, man-made equals dangerous' view of foods is the one put forward by Bruce Ames and colleagues. Fundamental to the safety assessment of any

potentially toxic substance is the maxim attributed to Paracelsus, that the effects on the body of any substance, good or bad, depend on the dose. Ames pointed out that if the same precautionary criteria that are used to set pesticide safety levels — toxicological data, including tests on rodents for carcinogenicity — were applied to the natural toxins in plants that have evolved to deter predators, many foods would be deemed unsafe. For example, potatoes, grilled food and peanuts would be banned if they underwent the same kind of scrutiny as pesticide residues.

According to Ames, half of the natural toxins that have been tested (most have not) are rodent carcinogens, and each year the average American consumes about 10,000 times more of these natural pesticides than of synthetic residues. A single cup of coffee contains natural carcinogens equal at least to a year's worth of carcinogenic synthetic residues in the diet. The organic sector has claimed that its produce is lower in synthetic residues (fewer pesticides are used) but higher in natural toxins. From Ames's line of argument, consumers of organic produce may well be trading a minute amount of synthetic residue for equally — if not more — dangerous natural pesticides. This should, of course, be kept in perspective: any potentially detrimental effect of natural pesticides or synthetic pesticide residues is far outweighed by the health benefits of consuming five portions of fruit and vegetables per day.

Collman's quirky and erratic account, more a series of vignettes than a narrative, makes an effective case for not accepting the simple equation 'natural = safe'. In addition to food, he covers herbal medicines, environmental pollution, global warming, electromagnetic radiation and radioactivity. I would have liked a slightly less triumphalist tone, in recognition that there are still many uncertainties in our understanding of both environmental and diet-related impacts on human health. For instance, the toxicological consequences of exposure to cocktails of residues and the potential effects of long-term exposure are not well documented. As new data emerge, the experts, quite correctly, sometimes change their minds about safety limits. This recently happened for dioxins, for which the safety level has been reduced by a factor of five.

When viewed from the perspective of scientific uncertainty, some of the fears about unknown consequences may seem less irrational. A challenge for those responsible for translating science into regulatory policy is to find an effective way of taking people's concerns into account without straying from



Healthy debate: does natural really equal safe?

the bedrock of scientific evidence. There are no easy answers, but a start may be for scientists both to explain the uncertainties more fully, and to emphasize that evidence is dynamic and evolving rather than a set of ineluctable facts. ■

John Krebs is chairman of the UK Food Standards Agency and in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Life as a freeloader

Les associations du vivant:
L'art d'être parasite

by Claude Combes

Flammarion: 2001. 21 euros

J. C. Koella and C. D. M. Müller-Graf

Parasites are not only the scare of today's anti-terrorist squads, but also a masterful work of evolutionary art. The tiny mite *Histioglyphus laboratorum*, a parasite of *Drosophila*, launches itself as a miniature surface-to-air missile towards a fruitfly that is flying far above its head, an incredible feat of evolutionary engineering (the 'Star Wars' programme could learn a thing or two). Gravid mussels such as *Lampsilis ventricosa*,

TIM ROSKE/AP

when they are about to release their larval offspring from the gill tubes, display part of their mantle, waving and undulating excitedly, to imitate the little fish that are the main prey of predatory fish such as bass. Lured to the mussel with what seems to be a cheap and tasty meal, the greedy predator becomes the unfortunate host of the larvae, which are obligate parasites of fish.

'*The Art of Being a Parasite*' is an extensive collection of these and other wonderful and weird stories illustrating the various forms that parasitic life can take — and the elaborate ways of invading hosts and remaining in them — weaving the examples into a complex web to illuminate the ecology and evolution of interactions between species. It includes less-known examples, such as the only known cnidarian parasite (*Polypodium hydriforme*), which infects a developing egg of the beluga sturgeon (of course a French author would know the parasites of caviar) and goes on to develop into a stolon with miniature medusae, still within the egg, before being liberated into the water when the eggs are laid. Another gourmet example is the group of cestodes (*Amoebotaenia* spp.) in the gut of woodcock; this is the only bird eaten with its intestine, as the numerous cestodes growing within the gut give the bird its characteristic delicious taste (if you want to taste the worms while eating out in France, the bird is called a *bécasse*).

An extensive discussion focuses on the art of becoming — through evolutionary time — a parasite. The evolutionary progression of the parasitic way of life is particularly well illustrated with a group of prosobranch molluscs. This progression starts with a species called *Capulus*, a superficial ectoparasite of marine invertebrates that attaches itself to their outside wall to profit from the water flow created by its host. It proceeds through a series of more parasitic forms, which develop a more intimate association with their hosts (starfish, sea urchins and sea cucumbers) by invading their body cavities while maintaining a connection with the external environment, to a fully endoparasitic species that parasitizes sea cucumbers. Although this sequence is not a phylogenetic lineage, it is quite astonishing how well one example succeeds the other.

Although this is clearly a book that concentrates on parasites, Claude Combes does not forget to discuss mutualism as one end of a spectrum of species interactions, using among other examples the mutualistic bacteria–nematode complexes (such as *Heterorhabditis*–*Photorhabdus*) that parasitize insects. The nematode penetrates the insect, releases immunosuppressive substances into the insect's haemocoel, and then releases its bacteria. Uninhibited by the insect's immune response, these replicate extensively while producing antibiotics to keep out potential bacterial competitors. After several days the insect dies, and the



Tasteful: the woodcock is the only bird eaten with its intestine, as parasites in the gut give it its flavour.

nematode (which has benefited from the meal provided by the mix of the insect's tissue and the bacteria, and has reproduced) takes up and maintains a hundred or more bacteria, then leaves the insect to find a new victim.

Many other aspects of host–parasite relationships are discussed: host defence mechanisms of hosts, counter-mechanisms of parasites, manipulation of host behaviour, and the importance of parasites in ecosystems and for humans in particular. Although the author is at his best when he illustrates these ideas with the natural history of the host–parasite associations, he also attempts to introduce the evolutionary foundation underlying host–parasite coevolution to the general public — for whom this book seems to be written. Although this attempt is praiseworthy, the book is at its weakest here. The discussions are sometimes circular and are often treated too lightly to be of much help. Perhaps the most striking example is the role of parasites in the maintenance of sexual reproduction and genetic recombination: this extensive field of evolutionary research is discussed in only one paragraph. This criticism is, however, not to be taken too severely. Combes's book does not purport to be a detailed scientific discussion of

evolutionary parasitological processes; its strength is rather to illustrate the interactive forces involved in the coevolution of hosts and parasites with fascinating examples of their natural history. The book should be an asset to anyone teaching this subject.

On the practical side, an index would have made it easier to refer to earlier parts of the book and to find the interesting examples. For someone not familiar with the area, a glossary explaining some of the terminology would have been helpful, and could have alleviated some of the heavy footnotes.

In a nutshell, if you want to be introduced to the marvellous consequences of the evolution of parasites and their natural history, it would be difficult — despite some weaknesses in the evolutionary background — to find a more fascinating book. For the time being, however, you will have to read it in French. ■

J. C. Koella and C. D. M. Müller-Graf are at the Laboratoire de Parasitologie Évolutive, UMR 7103, Université Pierre et Marie Curie, 7 quai St Bernard, 75252 Paris, France.

Claude Combes's previous book *Parasitism: The Ecology and Evolution of Intimate Interactions*, has recently been published in an English translation (University of Chicago Press, \$55).

A Universal view

Higher Than Everest: An Adventurer's Guide to the Solar System

by Paul Hodge
Cambridge University Press: 2001. 256 pp.
£18.95, \$27.95

Solar System Evolution: A New Perspective. 2nd edition

by Stuart Ross Taylor
Cambridge University Press: 2001. 484 pp.
£60, \$90

John E. Chambers

The Solar System is a truly remarkable place, and well worth a visit. This, at least, is the shared conclusion of two new books on the subject. But there the similarities end. *Higher Than Everest* is a light-hearted adventurer's guide to our planet's immediate neighbourhood, whereas Stuart Ross Taylor's book is an altogether weightier tome with a loftier goal in mind — to provide a scientific history of the Solar System.

Higher Than Everest is the kind of book that will appeal to armchair alpinists everywhere, and one that sets the standard for twenty-second-century aficionados of extreme sports. The book opens with an account of a hypothetical expedition to climb Olympus Mons, the tallest volcano in the Solar System that dwarfs all of the 'seven summits' here on Earth. Although the ascent is not difficult from a climber's point of view, would-be Reinhold Messners face a number of hurdles en route before they can bag the summit.

Paul Hodge provides a detailed plan of campaign with a generous amount of time set aside for scientific exploration along the way. This approach sets the tone for the rest of the book. As he moves to Venus, Io, Miranda and beyond, Hodge seizes every opportunity to insert snippets of hard science into his description of each breath-taking piece of scenery. Jargon is kept to a minimum, and the emphasis is on the 'what' rather than the 'why', but Hodge nonetheless manages to address a number of issues at the cutting edge of planetary science. Compared with Mars, journeys to the outer Solar System are technically challenging and fraught with uncertainty. But despite this, Hodge leaves the reader feeling that one day they might just happen.

Since the publication of the first edition of *Solar System Evolution* a decade ago, planetary science has advanced by leaps and bounds. The Edgeworth–Kuiper belt has been exposed (literally), the Galileo mission and comet Shoemaker–Levy 9 have transformed our view of the jovian system, and extrasolar planets — those beyond the Solar

System — are now an established fact. Yet despite these advances, understanding the origins and modern characteristics of the Solar System remains a frustratingly difficult task. This fully updated survey of the field is a noble attempt to make sense of the conflicting data sets and large gaps in our current knowledge.

The book has a strong cosmochemical slant. This is understandable given that meteorites and lunar and terrestrial rock samples provide some of the strongest constraints we have on theories about the origin of the Universe. Still, it is disappointing that certain other important areas are glossed over or abbreviated in the rush to provide a unified theory for the Solar System. For example, it is currently not known how a protoplanetary disk of gas and dust gives rise to mountain-sized 'planetesimals' — bodies that are large enough to have appreciable gravitational fields that can aid further coagulation. The universally favoured planetesimal theory of planet formation is the main theory advocated in *Solar System Evolution*, yet its shakier foundations are never examined.

Another bone of contention concerns the origin and structure of the giant planets. Some readers may not share Ross Taylor's conviction that the Sun's protoplanetary nebula was driven away by strong solar winds within the first million years of its existence. This is hard to reconcile with the standard model for the formation of giant planets, heartily endorsed here, which begins with the accretion of a ten-Earth-mass solid core, followed by the gradual acquisition of a gaseous envelope. The alternative theory, that Jupiter formed as a result of the gravitational collapse of part of the Sun's protoplanetary disk, is given short shrift, although the evidence for a jovian core is relatively weak.

The book opens with a sombre account of the achievements of the pioneers of planetary science. This is followed by a dizzying ride through galaxy formation, nucleosynthesis — the formation of atoms through nuclear reactions in the cores of stars — and the origin of stars. We then finally settle down to the serious business of contemplating the Solar System. On the whole, Ross Taylor's writing is fluent and engaging, but he occasionally slips into the annoying habit of using specialized terminology before defining it.

The second half of the book contains a piece-by-piece analysis of each of the main components of the modern Solar System. There is an excellent comparative analysis of Venus, Earth and Mars, which includes a detailed discussion of why Earth alone possessed the conditions necessary to spawn plate tectonics and the formation of a continental crust. In the outer Solar System, the author boldly embraces the idea that Pluto is not a major planet, but rather the

leading member of a new breed. The recent discovery of an object in the Edgeworth–Kuiper belt that has a diameter roughly half that of Pluto's provides timely support for this viewpoint.

A central theme of the book is that the Solar System is unique, the result of a long series of chance events which are unlikely to be repeated elsewhere. Here the author is on firm ground. One of the most striking lessons we have learned from the discovery of extrasolar planets is that nature prefers variety to consistency when it comes to building planetary systems. Close analogues of the Solar System may turn out to be rare, as are planets that resemble Earth, and this makes our own system all the more fascinating. Despite its occasional shortcomings, the scope of this study is impressive, and this well-researched book deserves a place on the bookshelf of anyone interested in the Solar System and its origins. ■

John E. Chambers is at the NASA Ames Research Center, MS 245-3, Moffett Field, California 94035, USA.



A jaunt through the Solar System

As its name suggests, *The Cambridge Photographic Guide to the Planets* by F. W. Taylor (Cambridge University Press, \$49.95, £29.95) offers a selection of the latest images of the planetary bodies in our Solar System. Starting with Mercury, the book takes the reader on a voyage to the outer edges of our planetary system, encompassing the planets and their respective satellites, plus asteroids and comets along the way. The image shown above, for example, shows Jupiter and was taken by the Hubble Space Telescope. It offers a relatively rare 'true colour' picture of the planet — in most images of Jupiter, the colours are exaggerated to emphasize the contrasts.

Talking techno

Language and the Internet

by David Crystal

Cambridge University Press: 2001. 272 pp.

£13.95, \$19.95

Geoffrey Nunberg

It's hard to imagine any technology surpassing print in its linguistic effects, such as the fixing of spelling and grammar and the creation of standardized national languages that underlie the formation of the modern nation states. Yet the effects of digital technology could be even more far-reaching. After all, the printing press was merely a new means of production, and the 'print revolution' actually depended on a number of later developments as well, such as cheap pulp paper and steam-powered roller presses — not to mention ancillary technologies such as the telegraph and the railroad. The networked computer, by comparison, has worked instant transformations not just in the way documents are produced, but in the way they are composed, disseminated and read. In his engaging and provocative book *Language and the Internet*, the British linguist David Crystal argues that networked computers are already giving rise to a new form of language — Netspeak — characterized by both its vocabulary and its graphical devices, and by communicative conventions that have no analogues in predigital communication.

True, there is a certain overgeneralization in talking about the Internet's language as a single variety. (By analogy, imagine someone coining the term 'printspeak' to refer to a single language variety that covers everything from the articles in *Nature* to the features in the *Sun* or *The New York Post*, not to mention the language of handbills, lawn-mower manuals and rock-concert posters.) And Crystal acknowledges the heterogeneity of Netspeak in separate chapters on the language of chatgroups, e-mail, the web and the 'virtual worlds' called MUDs or Moos, where users enact fantasy roles.

Still, *Netspeak* has some common features that set it apart from other forms of speech. Granted, many of these are marginal or ephemeral: the hacker slang (such as 'bogus' and 'dubiosity'); the coy acronyms (AFAIK for "as far as I know", FWIW for "for what it's worth"); and the 'smileys' — strings of punctuation marks such as :-) and :-(that e-mail writers use to indicate smiles, winks, frowns and the like. (You wonder what Jane Austen would have made of a device that relieved the reader of the need to guess about an ironic intent.) Lexical and orthographic antics such as these have a certain charm, but they are unlikely to survive much longer than the slang of the US CB radio craze of the 1980s.

But other features of Netspeak are likely to be more enduring. As Crystal points out,

e-mail imposes a certain brevity on communication, as most people are reluctant to scroll down to read more than a screenful of material. It makes possible a new style of intercalated text and response, as 'indents' pile up in the left-hand margin. And it encourages many writers to dispense with the formalities of a traditional letter: I routinely receive communications from people I don't know that begin with a bare "Hi" and close breezily with "Cheers", an item that North Americans used to regard as strictly an anglicism. (To be sure, this informality may be more marked in the Anglo-Saxon world; even in e-mail, some French correspondents still insist on florid valedictions such as "I dare to hope you will do me the honour of a response".) More than anything else, what sets Netspeak apart is this informality of style, and the new forms of discourse it heralds. As Crystal observes, the situation of Netspeak has as much in common with speech as with traditional writing — it is time-governed and transient, and its utterances "display much of the urgency and energetic force that is characteristic of face-to-face conversation".

As these forms of communication become important elements of public life, they will inevitably shape both the way we use language and the common discourse we carry out in it. For some, the Internet opens up new possibilities for direct democracy, as citizens become 'netizens' exchanging their views in the new 'electronic commons'. But this informality can also have the effect of disenfranchizing some people, as it

substitutes for the neutral public language of print a new language that leans more heavily on the implicit norms of the conversation of the Anglo-US middle classes. (A French biologist I know who has published widely in English confessed himself to be at a loss to follow the give-and-take of the conversations on scientific discussion groups on the Internet — "Ce n'est pas l'anglais que nous employons à la Sorbonne!")

But the Internet has already demonstrated a remarkable capacity to surprise us. Just a few years ago, many people thought that the Internet would be a path for the universal triumph of English, a view that Crystal himself endorsed in his earlier book *English as a Global Language* (Cambridge University Press, 1997). But by now, he acknowledges, the early discrepancies in language use are evening out. The proportions of web pages in languages such as French, Spanish, German and Japanese are increasing rapidly, and even relatively minor languages are achieving a foothold on the web — not surprising, when you consider that the costs of entry are far lower than they are for print.

In the end, the one thing we can be sure of is that the Internet will be a home for lively chatter, whatever it sounds like. As Crystal concludes, "The arrival of Netspeak is showing us *homo loquens* at its best". ■

Geoffrey Nunberg is at the Center for the Study of Language and Information, Stanford University, Ventura Hall, Stanford, California 94305-4115, USA. His most recent book is *The Way We Talk Now* (Houghton Mifflin, 2001).

Gone — but not forgotten



Agriculture in the United States has changed dramatically since the 1940s. *Changing Works: Visions of a Lost Agriculture* by Douglas Harper (University of Chicago Press, \$35, £22.50), documents the shift from small units and a cooperative approach to the industrialized, mass-production systems seen today. Drawing on a host of photographs from the 1940s, the book also illustrates the unstoppable rise of mechanization and automation in the industry.



Seeking universals

Melvin Konner

Despite the successes of behavioural biology, some still deny that human nature exists. It is difficult to understand what this denial means. Surely such features as, say, bipedal walking, language and a dependence on eight amino acids are integral to our being? Not necessarily. People who cannot walk or talk are still human. Phenylalanine is poisonous to some people. Many other species need those amino acids. And the enormous variety of human languages would seem to be the final blow.

Actually these answers, valid in themselves, are irrelevant, little more than a debater's move. The debate becomes especially fierce and the evasion especially deft when behaviour is discussed, particularly if the behaviour is deemed undesirable — gluttony, lust or violence. The real issue is easily drawn: are some things in human behaviour strongly and inherently part of what we are? As with nucleic-acid sequences and skeletal morphology, some features vary widely within the species, some little or not at all.

The genome sequencers have identified features that we all share. Only 1 or 2% are unique to us, but all are part of our nature.

Characterizing human behaviour, emotion and mind can be much less expensive and at least equally successful. As with the genome or the skeleton, neither the absence of perfect invariance nor an extension to other species prevents a feature from helping us to define our humanity. In fact, a behaviour can be far from universal and still shed light on our nature. As Joseph Greenberg pointed out, any feature of phonology or syntax that is not distributed randomly over the world's languages must be guided by a bias, which needs explaining. The bias is in the nature of language and is therefore in human nature.

Not all of human behaviour lends itself to such rigorous analysis as phonology and syntax, but it can be studied. The search for universals is a good initial step. Where variation ends, human nature begins; where variation is systematic rather than random, the biases lead to laws of human nature. However, even 'universal' needs parsing. It can mean that every normal member of the species has it, as with bipedal walking or language. But it can also mean that every member of an age or sex class exhibits it, as with the Moro reflex of newborn infants or the ejaculatory sequence of post-pubertal males. It can mean just that some members of every population show it; thus, paradoxically, it is universal despite being unusual, as with depression or homicide. It can be cultural, as with incest prohibitions, marriage or the mastery of fire. And finally, far from being rigid, what is universal may be only the pattern of variation, as in the three- or five-factor models of personality.

Where do such biases come from? Ultimately they come from the genome, through molecular function and anatomical structure, especially in the neural and endocrine systems. But many variables can intervene, and we must be careful about unknowns in the causal chain. The division of labour by sex once seemed to be universal; now it is largely superseded by cultural and technological development. The human genome generates a division of labour by sex in certain circumstances: high infant and child mortality; limitations of suitable weaning foods; the need for women to be pregnant or lactating; and the value of physical strength for hunting, large-animal husbandry and pre-modern war.

But change those circumstances and the 'universal' disappears. Consider smiling in greeting, a feature shown by all normal humans over eight weeks old — or, more precisely, over forty-eight weeks of conceptual age. This stronger link to conceptual than postnatal age strengthens the case for smiling in greeting as a feature of human nature, because it suggests a genetically guided process of neural development rather than a

Human nature

In the era of genomics and brain imaging, hypotheses about human nature are more testable than ever.

specified dose of experience. In fact, experience may have a modest role; socially deprived or blind infants smile with small delays. In adult neurology patients, the loss of smiling on command stems from a motor-cortex lesion, and that of spontaneous smiling from a subcortical one. Moreover, in Japan and England smiles are more regulated than in Latin America. But cultures everywhere show spontaneous smiling in greeting. Likewise, if I say that grieving the loss of a loved one is part of human nature, I refer to similar evidence. Neither individual exceptions nor cultural rules of expression undermine the hypothesis that it is human nature to grieve.

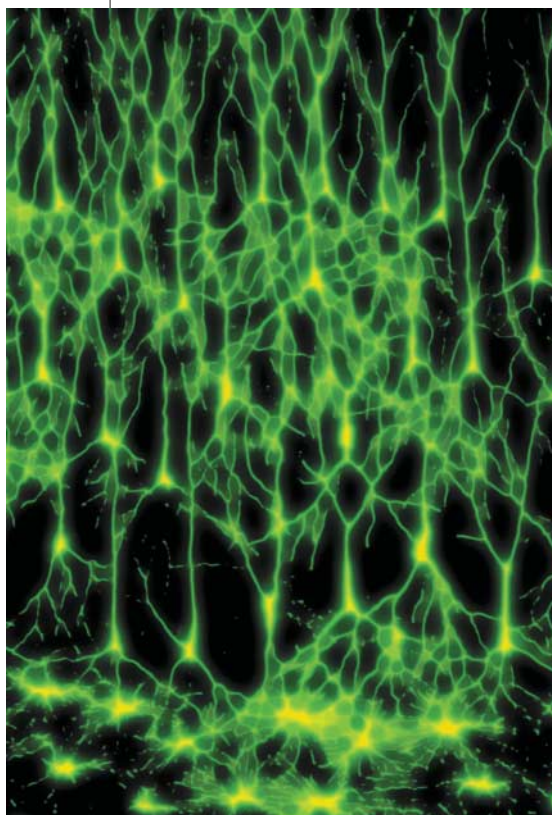
Paul Griffiths redefines emotions as phylogenetically shared affect programs. This makes them likely to be genetically guided nervous-system functions, beyond arousal or learning. Darwin's example of this, blushing, is unique to humans but physiologically biased and phylogenetically based. Some cognitive features of mind could be similarly defined; others wouldn't stand the test of shared phylogeny. This is true of emotion too.

It would be disingenuous to omit hatred, lust and greed, which come to mind most readily when human nature is mentioned. Why can't we calmly try to find out just how deeply ingrained they are? In the era of genomics and brain imaging, hypotheses about human nature are more testable than ever. Must we take offence at every potential limit that biology may place upon our freedom? Or can we take comfort that there are some that we all share? Only if human nature exists can we truly say that to gaze at a stranger's face is like looking in a mirror, and one that reflects things behind the face. ■

Melvin Konner, who teaches at Emory University, Atlanta, Georgia, USA, is the author of the revised edition of *The Tangled Wing: Biological Constraints on the Human Spirit*.

FURTHER READING

Darwin, C. *The Expression of the Emotions in Man and Animals* 3rd edn (Oxford Univ. Press, Oxford, 1998).
Eibl-Eibesfeldt, I. *Human Ethology* (Aldine de Gruyter, New York, 1989).
Ekman, P. & Rosenberg, E. L. (eds) *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System* (Oxford Univ. Press, Oxford, 1997).
Griffiths, P. *What Emotions Really Are: The Problem of Psychological Categories* (Univ. Chicago Press, Chicago, 1997).
Greenberg, J. et al. (eds) *Universals of Human Language* (Stanford Univ. Press, Stanford, 1978).



Innate nature: are some so-called cultural phenomena wired in the brain, even before birth?

Protein complexes take the bait

Anuj Kumar and Michael Snyder

Many cellular functions are carried out by proteins that are bound together in complexes. In two new large-scale studies, labelled proteins are used as 'bait' to capture and identify those complexes.

To appropriate a quote from John Donne, "no protein is an island entire of itself" — or at least, very few proteins are. Most seem to function within complicated cellular pathways, interacting with other proteins either in pairs or as components of larger complexes. A comprehensive understanding of these interactions will be needed before we can appreciate the mechanisms by which cellular pathways function and interlink. On pages 141 and 180 of this issue, Gavin *et al.*¹ and Ho *et al.*² describe significant advances towards this goal. Each group has characterized hundreds of distinct multiprotein complexes in the budding yeast *Saccharomyces cerevisiae*, using approaches in which individual proteins are tagged and used to pull down associated proteins, which are then analysed by mass spectrometry.

These studies^{1,2} exemplify an emerging paradigm in protein biology: the systematic analysis of an organism's complete complement of proteins (its 'proteome'). Protein interactions on a proteome-wide scale have already been analysed in several ways. In a pair of landmark papers, Uetz *et al.*³ and Ito *et al.*⁴ adapted the yeast 'two-hybrid' assay — a means of assessing whether two single proteins interact — into a high-throughput method of mapping pair-wise protein interactions on a large scale. The authors collectively identified over 4,000 protein–protein interactions in *S. cerevisiae*. Our own group⁵ has developed a microarray technology in which purified, active proteins from almost the entire yeast proteome are printed onto a microscope slide at high density, such that thousands of protein interactions (and other protein functions) can be assayed simultaneously.

Gavin *et al.*¹ and Ho *et al.*² take a different approach — one that is particularly effective at identifying protein complexes that contain three or more components. Large-scale efforts to characterize protein complexes are generally rate-limited by the need for a nearly pure preparation of each complex. In the new studies^{1,2}, protein complexes were purified as follows (Fig. 1). First, the authors attached tags to hundreds of different proteins (to create 'bait' proteins). They then introduced DNA encoding these bait proteins into yeast cells, allowing the modified proteins to be expressed in the cells and to form physiological complexes with other proteins. Then, using the tag, each bait pro-

tein was pulled out, often fishing out the entire complex with it (hence the term 'bait'). The proteins extracted with the tagged bait were identified using standard mass-spectrometry methods.

Applying this approach on a proteome-

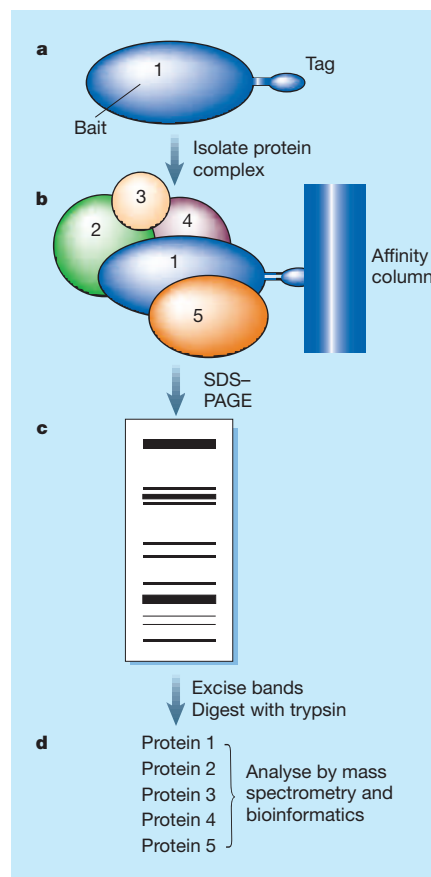


Figure 1 Analysing protein interactions. In the 'co-precipitation/mass spectrometry' approach used by Gavin *et al.*¹ and Ho *et al.*², an 'affinity tag' is first attached to a target protein (the 'bait'; a). b, Bait proteins are systematically precipitated, along with any associated proteins, on an 'affinity column'. c, Purified protein complexes are resolved by one-dimensional SDS-PAGE, a technique that involves running an electric charge through the complexes on a gel, so that proteins become separated according to mass. d, Proteins are excised from the gel, digested with the enzyme trypsin, and analysed by mass spectrometry. Database-search algorithms (bioinformatics) are then used to identify specific proteins from their mass spectra.

wide scale, Gavin *et al.*¹ have identified 1,440 distinct proteins within 232 multiprotein complexes in yeast. As 91% of these complexes contain at least one protein of previously unknown function, the study provides a wealth of new information on 231 previously uncharacterized yeast proteins, and on a further 113 proteins to which the authors ascribe a previously unknown cellular role. Furthermore, Gavin *et al.* find that most of these complexes have a component in common with at least one other multiprotein assembly, suggesting a means of coordinating cellular functions into a higher-order network of interacting protein complexes.

An understanding of this high-order organization will undoubtedly offer insight into corresponding networks in other organisms, as most yeast complexes have counterparts in more complex species (one reason why researchers are interested in this unicellular organism). Gavin and colleagues illustrate this point by purifying and analysing three equivalent multiprotein complexes from yeast and human cells: the Arp2/3 complex, a component of the cellular 'skeleton'; the Ccr4–Not1 complex, which is found in the nucleus; and the TRAPP complex, which is involved in transport from one intracellular compartment (the endoplasmic reticulum) to another (the Golgi). In each case, the authors retrieved human and yeast complexes that were similar, if not identical, in composition.

Using the same general approach, Ho *et al.*² constructed an initial set of 725 yeast bait proteins, from which they identified 3,617 interactions involving 1,578 different proteins. They describe interaction networks assembled around the protein kinase Kss1 — a known component of pathways involved in mating and filamentous growth — and complexes associated with the cyclin-dependent kinase Cdc28 and the gene-transcription factors Fkh1 and Fkh2. In addition, Ho and colleagues used 86 bait proteins that are implicated in the DNA-damage response, allowing them to delineate much of the yeast damage-response network. In particular, they reveal many regulators and targets of the protein kinase Dun1, and a possible role for the DNA-repair protein Rad7 in processes of targeted protein degradation.

The approach taken by Gavin *et al.* and Ho *et al.* is clearly powerful, but it does have

drawbacks. Both groups find a significant number of false-positive interactions, while failing to identify many known associations. Gavin *et al.* estimate that 30% of the interactions they detect may be spurious, as inferred from duplicate analyses of 13 purified complexes. Conversely, they failed to detect any interacting partners for Bmh2 (ref. 6), a regulatory protein that has previously been shown to interact with a number of other proteins, including Ste20 (involved, for example, in yeast mating)⁷, and Msn2 and Msn4 (stress-responsive transcription factors)⁸. Ho *et al.*, meanwhile, did not detect nucleotide excision repair factor-2, a tight complex⁹ that contains the well-characterized DNA-repair proteins Rad4 and Rad23. So, as in most large-scale studies, these results are imperfect. It will be essential to integrate data from many different sources to obtain an accurate understanding of protein networks.

Proteomic studies such as these^{1,2} have generated a huge volume of exciting data. Yet — setting aside the problem of false positives and negatives — there is much still to be learned before we have a comprehensive knowledge of functional pathways within even a model organism such as yeast. To understand the magnitude of the task, consider the yeast proteome. Assuming that each

protein interacts with an average of five partners — a reasonable estimate drawn from experience and preliminary two-hybrid results — the yeast proteome should encompass some 30,000 protein interactions, many of which change during the life cycle of the organism. So far, protein microarray analyses and studies like those of Gavin *et al.* and Ho *et al.* have collectively identified, at most, 11,000 different protein associations (and probably fewer, considering the potential overlap between data sets). Although feasible, the characterization of all remaining interactions will almost certainly be labour intensive. But the resulting data will be more than worth the effort. ■

Anuj Kumar and Michael Snyder are in the Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06520-8103, USA.
e-mails: anuj.kumar@yale.edu
michael.snyder@yale.edu

1. Gavin, A.-C. *et al.* *Nature* **415**, 141–147 (2002).
2. Ho, Y. *et al.* *Nature* **415**, 180–183 (2002).
3. Uetz, P. *et al.* *Nature* **403**, 623–627 (2000).
4. Ito, T. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
5. Zhu, H. *et al.* *Science* **293**, 2101–2105 (2001).
6. Gelperin, D. *et al.* *Proc. Natl Acad. Sci. USA* **92**, 11539–11543 (1995).
7. Roberts, R. L. *et al.* *Cell* **89**, 1055–1065 (1997).
8. Beck, T. & Hall, M. N. *Nature* **402**, 689–692 (1999).
9. Guzder, S. N., Sung, P., Prakash, L. & Prakash, S. *J. Biol. Chem.* **273**, 31541–31546 (1998).

Oceanography

Bubbling under

Chris German

The study of hydrothermal vents is a young and fertile discipline. The latest findings, and the enticing prospects offered by new technology, came in for discussion at two meetings held late last year.

Deep-sea hydrothermal activity was discovered only in 1977, on the sea floor near the Galapagos Islands, so it is little wonder that the phenomenon continues to spring surprises. Hydrothermal vents arise where cold sea water interacts with freshly formed, hot ocean crust along chains of submarine volcanoes, termed mid-ocean ridges, which run for more than 50,000 km across the world's ocean basins. Two different meetings on hydrothermal systems were held late last year, in San Francisco and Berlin*, attesting to the vigour of research on the topic.

The event in San Francisco was convened in honour of the geochemist John Edmond, who died last year aged 57. Edmond was the first to recognize the full importance of hydrothermal activity. He used the first

observations from near the Galapagos Islands to make prescient estimates about the extremely high temperatures (around 350 °C) that hydrothermal fluids can reach, and discussed their significance in terms of chemical fluxes to the oceans ("These," he remarked laconically in a 1979 paper¹, "are large.") Fittingly, at San Francisco, it was two of Edmond's former students who described the most exciting new results.

Beneath the permanent ice cover of the Arctic Ocean, evidence has been found of extensive hydrothermal activity along the world's slowest-spreading and least volcanically active plate-tectonic boundary, the Gakkel Ridge (H. Edmonds, Univ. Texas). This discovery is as notable for astrobiologists as for others, in that it will provide a natural testing ground for techniques that could be used to explore Jupiter's ice-bound satellite Europa. Meanwhile, preliminary investigations of some of the first vent fluids to be collected from the Indian ocean — from one of only two known sites, both identified

within the past year or so — have shown that these samples are more chemically enriched than similar fluids studied previously (K. Von Damm, Univ. New Hampshire).

Soon after seafloor venting was first discovered, it was proposed that hydrothermal activity may be geographically rare and occur only along the volcanically active East Pacific Rise. But the news about Gakkel Ridge adds to a mounting body of evidence that hydrothermal activity can occur no matter how low the rate at which magma is supplied from the underlying mantle. As recently as ten years ago, it was commonly thought that hydrothermal fluids should all have a similar composition that would remain invariant for decades. On the evidence of Von Damm's report, however, it seems that we are still some way from understanding what the full flux from hydrothermal activity to the oceans may be. Clearly, the world is not nearly as straightforward as we once thought.

But why study these hydrothermal systems? Surely, even if they are dramatic (Fig. 1), they are of only marginal significance and appeal? The answer has several facets, not least of which is that hydrothermal activity is part of the way in which the Earth works. Such activity is intimately linked to plate tectonics and occurs almost exclusively at ocean ridges, yet to Earth scientists it represents a significant cooling process for the planet as a whole. For oceanographers, hydrothermal fluxes of chemicals entering the oceans (notably iron and manganese, but also methane, hydrogen and hydrogen sulphide) are comparable to those from rivers, and help to buffer the chemical balance of the oceans and atmosphere. The best estimates suggest that the entire volume of the oceans is processed through high-temperature vents every ten million years, and through deep-sea hydrothermal plumes every few thousand years. The latter timescale is similar to that of the mixing time for the modern-day deep ocean, and shorter than that of glacial-interglacial oscillations.

Hydrothermal vents offer plenty for life scientists, too: more than 400 new species have so far been found at vent sites, representing the discovery of one new species every two weeks, on average, throughout the past quarter-century. These organisms range from anaerobic microbes all the way to spectacular red-plumed tubeworms, which can reach several metres in length. In between are organisms such as the blind shrimps of the Mid-Atlantic Ridge, now also known to be present in the Indian Ocean, which are only a few centimetres long but swarm in millions around individual vent sites. Hydrothermal vent microbes seem to sit at the very base of the 'tree of life', and provide insights into both the origins of life here on Earth and its possible existence elsewhere in the Solar System.

*Pushing the Envelope: A Tribute to the Career and Accomplishments of John M. Edmond, American Geophysical Union meeting, San Francisco, 11–15 December 2001; Energy and Mass Transfer in Marine Hydrothermal Systems, 89th Dahlem Workshop, Berlin, 15–19 October 2001.



Figure 1 Deep-sea drama — a submersible hovers close to a 'black smoker' hydrothermal vent. This vent is on the East Pacific Rise, some 2,500 metres below the ocean surface.

Clearly, research on hydrothermal vents has to draw on a range of expertise. It was in recognition of this multidisciplinary requirement that the conference in Berlin was convened, with the main aim of discussing what we don't know, but may yet find out.

One theme was the extremes of temperature, pressure, pH and salinity under which hydrothermal fauna thrive, and what this can tell us about the limits to life on Earth. A typical hydrothermal vent may emit fluids that are not only hot (350 °C or more), but also acidic (pH 3–4) and highly toxic, with high concentrations of dissolved hydrogen sulphide and heavy metals such as cadmium, arsenic and lead. Indeed, life on our planet is the rule rather than the exception — it has adapted to almost the entire range of physicochemical variables observed, at least on Earth's outermost shell (thinking laterally, one can conclude that the most hostile environments for microbes are not 'natural' but those that are policed by the food-hygiene industry). Even quite conservative estimates suggest that the inclusion of all the microbes that are likely to be present beneath seafloor hydrothermal systems could double estimates of the total biomass on Earth (M. Summit, Washington Univ., St Louis; J. Baross, Univ. Washington, Seattle).

Whenever volcanoes erupt on the sea floor, huge quantities of these microorganisms, together with life-sustaining chemical nutrients, are released into the overlying water column (J. Cowen, Univ. Hawaii). Deep-ocean currents then propel these materials along the ridge crest (K. Speer, Florida State Univ.). Taken together, these findings suggest that there is a corridor of high-productivity water just a few tens of kilometres wide, and less than 500 metres

from top to bottom, that threads its way through the deep ocean, allowing rapid colonization of new vent sites. Understanding the diversity and distribution of fauna at these different vent sites is one of the main threads of modern hydrothermal research². Understanding how hydrothermal systems evolve — physically and chemically, over years and decades — is an equally tantalizing goal.

Behavioural science

Homo reciprocans

Samuel Bowles and Herbert Gintis

Humans are often generous, but cooperation unravels when others take advantage of them. Many people punish such 'free riders', even if they do not benefit personally, and this 'altruistic punishment' sustains cooperation.

Garrett Hardin¹ famously described a group of herders whose pursuit of self-interest leads to overgrazing of a pasture, driving it to ruin. His term for the process, the 'tragedy of the commons', underlined its inexorable nature. But herders, fishers and other users of common resources frequently avert the tragedy, devising ways to pursue common objectives and to curb free riding². Indeed, it is a distinctive characteristic of humans that we cooperate in large groups of people to whom we are not related — giving to charity, participating in political movements, and conforming to social norms. On page 137 of this issue, Fehr and Gächter³ describe the outcome of an experimental game involving human players which suggests that a key ingredient in human cooperation may be the punishment of free riders by their more public-spirited peers.

news and views

The way forward is in developing deep-sea observatories for long-term interactive investigation, using Internet connections made feasible with the advent of fibre-optic cables and global satellite communications. A necessary step in this direction was the announcement in San Francisco that new JASON-II robot submersibles are to be built, one each for Britain and the United States^{3,4}. These vehicles have advanced manipulator dexterity and will be able to carry payloads of up to 100 kg. Unlike manned submersibles, shifts of pilots, engineers and scientists can be changed without the robots ever leaving the sea bed, so maximizing their efficiency.

The announcement coincided with preview screenings of an IMAX movie, *Voyage to the Abyss*, set for general release in late 2002. This provided a glimpse of what both scientist and layperson can expect to experience in the future: interactivity with a deep-sea environment that is being observed remotely but in real time, with the capacity to reset experiments, relocate sensors and so on. And all this without enduring the deprivations of shipboard life.

Chris German is at the Southampton Oceanography Centre, European Way, Southampton SO14 3ZH, UK.

e-mail: cge@soc.soton.ac.uk

1. Edmond, J. et al. *Earth Planet. Sci. Lett.* **46**, 1–18 (1979).

2. <http://www.wm.edu/com1>

3. <http://www.soc.soton.ac.uk/OED/ROV/index.php>

4. <http://www.marine.who.edu/ships/rovs/upgrades.htm>

Box 1 'Public goods' experiments

In the 'public goods' game, each subject is given an endowment, w . Each subject may contribute any part of this endowment to a common account. Following this contribution stage, there is a pay-off stage in which the experimenter gives each of n subjects a fraction, $q \in (1/n, 1)$, of the amount in the common account. Contributing is an altruistic act as it increases the average group pay-off (because $q > 1/n$) at a cost to oneself (because $q < 1$).

In the 'public goods with punishment' game, following

the contribution stage there is a punishment stage in which the contributions of each subject, a_i , are made public and each player is given the opportunity to reduce the pay-offs of subject j by m_j at a cost to the punisher of cm_j . After the punishment stage, the pay-off stage is as above. Thus the pay-offs to subject i are: $y_i = w - a_i + q\sum a_j - \sum m_j - c\sum m_j$ where the summation is over all the group members. What the subject keeps is captured by the first two terms ($w - a_i$); $q\sum a_j$ is the amount received from the common account;

$\sum m_j$ is the cost of being punished by others; and $c\sum m_j$ is the cost of punishing others.

Interactions are anonymous and may last for a just a single round or a known number of rounds. In the 'partner treatment', membership of the groups remains unchanged during the game. In the 'stranger treatment', membership is shuffled after each round. The 'perfect stranger' treatment implemented by Fehr and Gächter³ ensures that no two subjects will interact more than once. **S.B. & H.G.**

this type of situation is just as altruistic as actually contributing to the public good. The key difference is that the punishment is conditioned by the behaviour of others — hence our term *Homo reciprocans*. Cooperation can be sustained because altruistic punishers induce even selfish group members to contribute to the common project. But is such altruistic punishment common?

Fehr and Gächter³ show that it is. The results of their well-designed experimental 'public-goods' game (Box 1) reveal that most people incur costs to themselves in order to punish those who have shirked in contributing to a public good. This form of altruistic punishment is common even when it is unmistakably the case that there can be no subsequent (direct or indirect) material benefits to those doing the punishing, because the groups of players are shuffled after every round such that no participant encounters any other more than once. Those who are punished for shirking contribute more in subsequent rounds, with the result that high overall levels of cooperation are sustained.

This result is striking in view of the fact that, in public-goods experiments with no provision for punishing free riders, contributions are substantial in early rounds of the game but dwindle to virtually nothing in subsequent rounds⁷. By showing that cooperation can be sustained by altruistic punishment where altruistic contribution may fail, the experiment will direct attention to why humans are so willing to punish those who violate norms, rather than focusing on why humans are also (sometimes) unconditionally generous to strangers.

Fehr and Gächter's work goes beyond previous studies^{8,9} by reshuffling participants such that no subject interacted more than once with any other. This treatment precludes both reputation and repetition, the

mechanisms underlying direct and indirect reciprocity and other self-interested explanations. As a result of this 'perfect stranger' treatment, subjects knew that punishing others could not raise their own pay-offs even if the shirkers they punished contributed more in later rounds. Rather, punishment *per se* provided the motivation, not some consequence anticipated by the player.

But why would the idea of punishing shirkers be a motivating factor? In the one-shot public-goods game, not punishing and not contributing maximize pay-offs irrespective of what the other group members do. So the willingness of subjects to reduce their own pay-offs by punishing shirkers but not to reduce their pay-offs by contributing to the public good in the absence of punishment needs to be explained in terms other than self-interest. Fehr and Gächter advance the view that punishment of shirkers is not evidence of a general-purpose predisposition to contribute to the public good, but rather reflects a negative emotional reaction to free riding. They stress a particular motivation — the desire to punish those who violate norms — rather than the more general motivation to contribute to the well-being of others. But the players are public-spirited nonetheless, as they do not themselves benefit from doing the punishing.

Shirkers, too, appear to have an emotional response to being punished. In other experiments¹⁰, even when punishment takes the form of verbal criticism rather than a pay-off reduction, shirkers contribute more in subsequent rounds, suggesting that punishment may evoke emotions of shame in the free rider.

Fehr and Gächter's experiment³ has implications for the design of constitutions and policies. It suggests that the objective should be to provide opportunities for the public-spirited to punish free riders, rather



100 YEARS AGO

Under the entrance gate, in the gravel, I saw a light of a brilliant greenish-bluish tint; it moved forward, leaving behind a trail of light which, gradually separating, became a scattered mass of brilliant points. The leading light had the form of a living, curving thread. A lighted match soon showed what the scattered points of light in its trail were, a dozen or so of red ants pursuing the *Geophilus*; one was clinging to it, each ant shone like a spark in the gravel, the centipede had discharged its fluid over them. I picked up the centipede and dropped it into a tumbler, where it splashed out a mass of light. Hurriedly placing my hand over the tumbler to prevent the insect from escaping, I felt suddenly a strange prickly sensation such as is caused by a slight contact with electricity, so that I hastily removed my hand, calling to a friend who, placing her hand over the tumbler, felt the same thing... Defence seems certainly to be one of the uses of this secretion, attributed by some authors merely to purposes of attraction. Rose Haig Thomas
From *Nature* 9 January 1902.

50 YEARS AGO

The arrival at Plymouth on December 6 of the Royal Research Ship *Discovery II* marked the completion of a twenty-month voyage of oceanographical survey... The Southern Ocean is a continuous belt of deep water encircling a central land mass; in consequence the prevailing currents and water circulation, and the fauna and flora, are distributed in a simpler pattern than in other oceans... A detailed knowledge of the bottom topography of the ocean is essential to the full understanding of the water circulation, and during the recent voyage the *Discovery II* has supplemented her earlier depth charts by an enormous number of new echo-soundings, and has recorded many continuous bottom profiles, running the sounding machine for long periods. Information of the depths of bottom sediments was also obtained in a number of places by seismic echo soundings with 1¼-lb. explosive charges. Four long bottom cores were obtained in the Indian Ocean with a Kullenberg corer. South of Australia, the corer struck a hard substance — possibly a meteorite — on the bottom; the nose-piece was blocked with hard brownish-black material which had to be chipped out. It was handed to the Geological Department of the University of Western Australia for further examination.
From *Nature* 12 January 1952.

than to assume, as David Hume¹¹ advised two-and-a-half centuries ago, that “every man ought to be supposed to be a knave and to have no other end, in all of his actions, than his private interest”.

Indeed, studies of contemporary hunter-gatherers and other evidence suggest that altruistic punishment may have been common in mobile foraging bands during the first 100,000 years or so of the existence of modern humans¹². A plausible explanation of the evolutionary success of this strategy is that groups with a high fraction of altruistic punishers would have sustained cooperation more successfully than groups with fewer punishers, and so would have prevailed over them¹³. Fehr and Gächter's results suggest that explanations for cooperation based on indirect reciprocity, reciprocal altruism or other forms of self-interest are at best incomplete. ■

Samuel Bowles is at the Santa Fe Institute, Santa Fe, New Mexico 87501 and the University of

Massachusetts at Amherst, Massachusetts 01003, USA.

e-mail: bowles@santafe.edu

Herbert Gintis is at the University of Massachusetts at Amherst, Massachusetts 01003, USA.

e-mail: hgintis@mediaone.net

1. Hardin, G. *Science* **162**, 1243–1248 (1968).
2. Berkes, F. *Nature* **340**, 91–93 (1989).
3. Fehr, E. & Gächter, S. *Nature* **415**, 137–140 (2002).
4. Axelrod, R. & Hamilton, W. D. *Science* **211**, 1390–1396 (1981).
5. Nowak, M. & Sigmund, K. *Nature* **393**, 573–577 (1998).
6. Gintis, H., Smith, E. A. & Bowles, S. *J. Theor. Biol.* **213**, 103–119 (2002).
7. Ledyard, J. O. in *The Handbook of Experimental Economics* (eds Roth, A. E. & Kagel, J.) 111–194 (Princeton Univ. Press, NJ, 1995).
8. Fehr, E. & Gächter, S. *J. Econ. Perspect.* **14**, 159–183 (2000).
9. Ostrom, E., Walker, J. & Gardner, R. *Am. Polit. Sci. Rev.* **86**, 404–417 (1992).
10. Bowles, S. & Gintis, H. in *Santa Fe Institute Working Paper* (January 2002).
11. Hume, D. *Essays: Moral, Political and Literary* (Longmans, Green (original publication 1754), London, 1898).
12. Boehm, C. *Hierarchy in the Forest* (Harvard Univ. Press, Cambridge, MA, 2000).
13. Bowles, S. *Economic Behavior and Institutions: An Evolutionary Approach to Microeconomics* (Princeton Univ. Press, NJ, 2002).

red, radio and X-ray wavelengths are largely unaffected. Indeed, infrared observations have revealed that several million stars are located within a light year of the centre. Within this central region is a strong radio source that is thought to be a supermassive black hole marking the true heart of our Galaxy. On page 148 of this issue, Wang, Gotthelf and Lang¹ report new observations from the Chandra X-ray observatory which provide the most detailed X-ray map yet of objects near the Galactic Centre.

At a distance of 26,000 light years from Earth, the centre of the Milky Way is the closest galactic nucleus available to study. A detailed picture of the physical processes that influence this extraordinary region is key to understanding all other galactic nuclei in the Universe. Improvements in the sensitivity and angular resolution of astrophysical instruments over the entire electromagnetic spectrum are currently revolutionizing our understanding of the Milky Way's nucleus. In particular, the results obtained by Wang and colleagues finally settle important aspects of the twenty-year-old debate surrounding the origin of X-ray emission from the Galactic Centre.

When X-rays were discovered at the Galactic Centre it was unclear whether the radiation was produced mainly by diffuse, hot interstellar gas, or by a combination of many individual sources. The unprecedented resolution of the Chandra measurements now allow astronomers to differentiate, in terms of both position and wavelength (spectrum), between emission from diffuse hot gas and point-like sources. The Chandra image on page 148 of this issue shows that at least 1,000 point-like X-ray sources contribute to the X-ray emission, only about 20 of which had been seen before. The authors argue that more than half of the newly discovered point sources are likely to be directly associated with the Galactic Centre, rather than with the extragalactic background. They also show that the combined spectrum of the point sources and the remaining diffuse component closely matches previous low-resolution measurements of the Galactic Centre, and so can explain most of the X-ray emission without requiring large amounts of hot gas.

Of crucial importance to Wang and colleagues' analysis are spectral features in the X-ray emission associated with the presence of iron in the Galactic Centre. There are two main components of this iron emission: an emission line at 6.7 keV, and a less energetic line at 6.4 keV (refs 2,3). Neutral iron has 26 electrons to compensate for the positive charge of its nucleus. The higher-energy 6.7-keV radiation occurs when an iron atom has lost all but two of its electrons and ends up in a highly ionized Fe²⁴⁺ state. The 6.4-keV line is caused by fluorescence of iron atoms weakly ionized by high-energy X-ray radiation or by collisions with cosmic ray particles.

Earlier low-resolution measurements of

Astronomy

X-rays reveal the Galaxy's centre

Andreas Eckart

For twenty years astronomers have wondered what is responsible for the X-ray emission from the centre of our Galaxy. New data from the sharpest X-ray eye around — the Chandra observatory — reveal all.

By the middle of the twentieth century astronomers knew that our Solar System was located far from the centre of the Milky Way — that broad swath of starlight you can see on a clear night. So what

is at the centre of our Galaxy? The Milky Way is some 80,000 light years across and contains more than a hundred billion stars. Clouds of dust prevent optical astronomers from seeing the Galactic Centre, but infra-

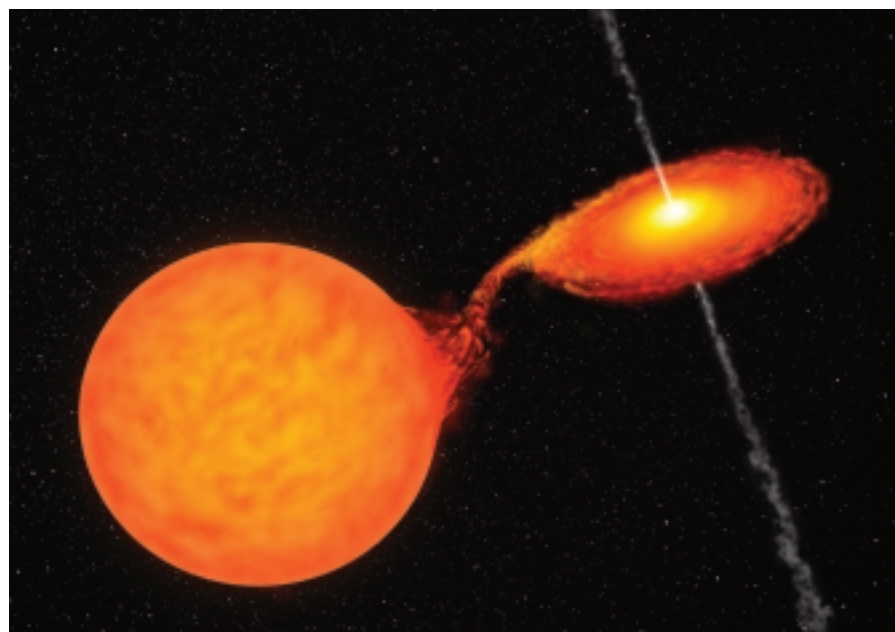


Figure 1 Computer illustration of an X-ray binary system. X-rays are emitted by the material streaming from the larger star to its compact companion, probably a black hole. According to new data from the Chandra X-ray observatory¹, these systems are responsible for most of the high-energy X-rays emitted by the Galactic Centre.

the 6.7-keV emission were attributed to a diffuse source of X-rays. This made it difficult to explain its origin because a diffuse heating source had to exist over a large volume. Now that most of this emission can be associated with individual point sources, a more straightforward physical explanation is possible: 6.7-keV emission is commonly produced by X-ray binary stars⁴. In these two-component systems, a dense object, such as a white dwarf, a neutron star or a stellar black hole, accretes matter from its less dense stellar companion, producing bright X-rays in the process (Fig. 1). Moreover, the 6.7-keV emission is highly characteristic of binary systems during the less active state in which they spend most of their lives, making this interpretation even more attractive.

The mystery of the 6.7-keV emission may have been solved, but the exact origin and nature of the remaining diffuse X-ray background is less clear. Most of the low-energy diffuse X-rays are likely to come from comparatively cool gas, typically found in young supernova remnants, and are therefore directly linked to the formation of massive stars⁵. Gas at even cooler temperatures cannot be observed in the Galactic Centre region because the view is obscured by dust clouds.

A more exciting possibility arises when we consider the origin of the diffuse X-rays associated with the fluorescent 6.4-keV emission. According to the Chandra data¹, nearly half of the diffuse emission in the range 5–8 keV is due to the 6.4-keV emission, but it is unclear what population of sufficiently bright and numerous X-ray sources could be causing this fluorescence. The authors suggest that the 6.4-keV line could be due to ancient X-ray emission⁶ from the supermassive black hole^{7–9} at the centre of the Galaxy.

Last year Chandra provided the first X-ray identification of this black hole by observing an X-ray outburst lasting just a few hours¹⁰. But it is possible that the central black hole is much less active now than it was during the past few hundred years. Higher activity in the past could explain much of the diffuse emission if energetic X-rays emitted by the black hole were scattered by colder surrounding material. Could this be a sign of a more dramatic past for the Galactic Centre?

Andreas Eckart is at the Institute of Physics, University of Cologne, Zùlpicher Strasse 7, 50937 Köln, Germany.

e-mail: eckart@ph1.uni-koeln.de

1. Wang, Q. D., Gotthelf, E. V. & Lang, C. C. *Nature* **415**, 148–150 (2002).
2. Valinia, A. *et al. Astrophys. J.* **543**, 733–739 (2000).
3. Tanaka, Y., Koyama, K., Maeda, Y. & Sonobe, T. *Publ. Astron. Soc. Jpn* **52**, L25–L30 (2000).
4. Barret, D. *et al. Astrophys. J.* **533**, 329–351 (2000).
5. Serabyn, E. & Morris, M. *Nature* **382**, 602–604 (1996).
6. Koyama, K. *et al. Publ. Astron. Soc. Jpn* **48**, 249–255 (1996).
7. Eckart, A. & Genzel, R. *Nature* **383**, 415–417 (1996).
8. Eckart, A. & Genzel, R. *Mon. Not. R. Astron. Soc.* **284**, 576–598 (1997).
9. Ghez, A. M., Klein, B. L., Morris, M. & Becklin, E. E. *Astrophys. J.* **509**, 678–686 (1998).
10. Baganoft, F. K. *et al. Nature* **413**, 45–48 (2001).

Molecular motors

Stretching the lever-arm theory

Michael A. Geeves

Motor proteins are essential to life: without them, all cellular transport would grind to a halt. New results on the size of steps taken by one family of motors, the myosins, will fuel the debate about how they move.

Most organisms, whether they consist of a single cell or billions, can move in a directed way, an ability that is largely attributed to molecular motor proteins. Of these, myosin II is perhaps the best understood, because of its role in muscle contraction. But other motors from the myosin family are also required for processes involving motility, from cell division to the transport of organelles within cells¹. One of the most hotly debated issues² in this field centres on how myosins move, and a theory known as the 'lever-arm hypothesis' has received much experimental support. This theory proposes that tiny changes in myosin's 'head' portion are amplified by the adjoining 'neck' (the lever arm) to produce large displacements at the far end of the neck that translate into movement of the whole protein³. The size of the displacement depends on the length of the lever (Fig. 1a).

Not everyone agrees, however, and the theory faces a new challenge from three studies — one of which is described on page 192 of this issue⁴ — of unconventional myosins. Tanaka *et al.*⁴ report that myosin V produces large, 35-nm steps regardless of the length of its neck. Meanwhile, Nishikawa *et al.*⁵ and

Rock *et al.*⁶ show that, despite its short neck, myosin VI also produces steps of 35 nm. The battle lines are clearly being drawn, with two other new studies^{7,8} of myosin V being interpreted in the classic lever-arm mode.

Myosins consist of a head, a neck and a tail. The neck comprises a structural element known as an α -helix, often attached to up to six polypeptide chains called light chains. The tail is involved in connecting the motor to its cargo, specifying the motor's cellular location and, in some cases, allowing dimerization. (Myosins II, V and VI, for example, all consist of two identical proteins, each with its own head, neck and tail.) The head attaches to, and moves along, tracks of actin filaments. These are made up of globular monomers that string together to form a chain; two chains twist round each other to form a helical filament.

The breakdown of the cell's energy store, ATP, powers myosin movement, driving large structural changes that cause the myosin heads to cyclically attach and detach from actin. Crystal structures of the myosin II head show the neck emerging from it, stabilized by two light chains, at an angle that varies by up to 60° depending on whether

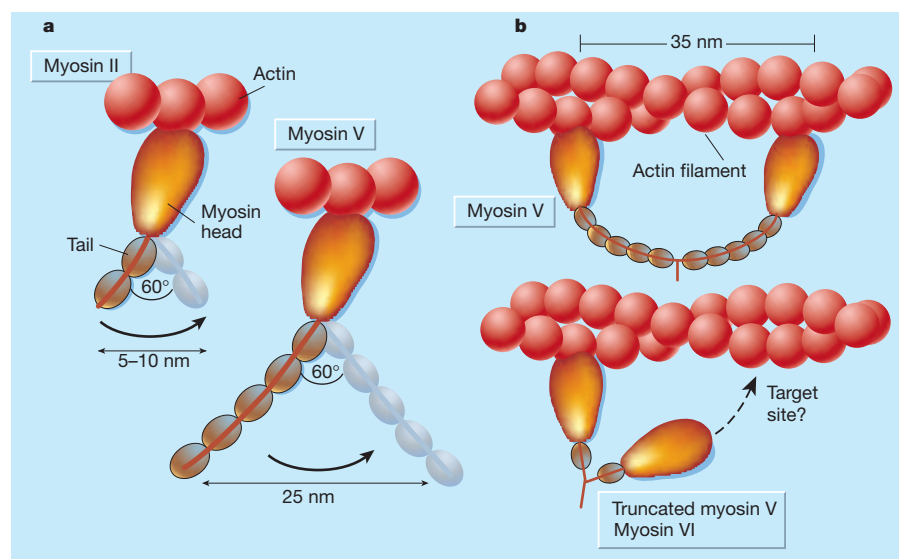


Figure 1 Motor movement. **a**, The length of the neck of a myosin motor determines the displacement of a cargo attached to it. The typical change in angle of the neck is 60°, so the neck displacement is theoretically 5–10 nm for a neck containing two light chains (such as myosin II) or up to 25 nm for a neck with five light chains (such as myosin V). **b**, The processive movement of a two-headed myosin motor. Top, the long neck of myosin V allows it to 'walk', hand-over-hand, taking 35-nm steps. Bottom, myosin VI has a short neck with just one light chain; truncated myosin V with a single light chain also has a short neck. Yet both can take long, processive steps of more than 20 nm^{4–6}. It is not known how.

the head is bound to ATP or to the products of ATP hydrolysis (ADP and phosphate). So the neck looks like a lever, which led to the idea that it operates as a rigid body to amplify small structural changes in the myosin head, with longer necks leading to a larger displacement. Support for this model comes from studies of myosins engineered to have lever arms of different length. The results show a linear relationship between the lever's length and the protein's speed of moving actin filaments over a surface^{9,10}, or step size in an optical trap^{8,11}.

Some electron-micrograph images of myosin V also support the lever-arm model¹². Most myosins, whether monomers or dimers, have short-lived interactions with actin and produce continuous movement only by working in groups. Myosins V and VI are unusual in that they are 'processive' dimers — they can take many hand-over-hand steps. But all myosins can bind only to certain points on actin, so they must find the next such site before committing to moving. Myosin V has a long neck with five light chains, and electron micrographs revealed that the neck allows the two heads to bridge not from actin monomer to monomer, a distance of 5.5 nm, but rather the longer, 35-nm pitch repeat¹². (So myosin V's processive movement should be linear, rather than following the twists of the actin helix — a big advantage if it's dragging an organelle behind it.) These images fit broadly with the lever-arm theory.

Toshio Yanagida's group, however, has long raised doubts about the model, and the most recent attack comes in the Tanaka *et al.* paper⁴. The authors engineered two myosin V constructs — one a native-like molecule with five light chains in the neck, and one with a single light chain and so a short neck. Remarkably, even for this truncated construct, the authors recorded long steps of 20–35 nm (Fig. 1b). The simplest interpretation is that the length of the neck does not define the step size for this myosin.

Nishikawa *et al.*⁵ (Yanagida's group again) and Rock *et al.*⁶ (James Spudich's group) present similar results for myosin VI — the only myosin known to travel backwards along actin¹³. Spudich's group has championed the lever-arm cause in the past, but both teams now find that myosin VI moves processively, taking steps of 20–40 nm along actin. Yet it naturally has a neck with a single light chain; the lever-arm model would predict that it takes steps of less than 10 nm.

These results^{4–6} suggest that, for myosins V and VI at least, the step size is defined more by the actin track than by the myosin itself. It was John Trinick who first used the analogy of walking along stepping stones: it does not matter what size step your legs would like to take, the only way to cross a lake and stay dry is to take steps defined by the distance between the stepping stones.

Of course, this only works if you can take steps of at least the distance between the stones. Two other new papers^{7,8} show that, for native myosin V, isolated single steps are shorter (at 20 nm) than processive steps, as is the first step of a processive series (matching the predictions of the lever-arm model). So a myosin's preferred step size may be shorter than the distance between actin sites. To account for the larger steps, the consensus appears to be moving towards a combination of the lever-arm swing orientating the stepping head towards the next available site on actin, with significant brownian motion then needed to locate and bind to that site. The different teams place different emphasis on the role of the lever^{7,8} and biased diffusion^{4,5}.

These mechanisms seem feasible when considering an unloaded motor, but a question for the future is whether they can work when the motors are carrying cargo. So far, only Spudich's group is using an optical trap with a feedback system to maintain constant

load on the motors over the large distances that they travel. As ever, the competition is as much about developing the technology as it is about interpreting the data. ■

Michael A. Geeves is in the Department of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, UK.

e-mail: m.a.geeves@ukc.ac.uk

1. Myosin Home Page; www.mrc-lmb.cam.ac.uk/myosin/myosin.html
2. Cyranoski, D. *Nature* **408**, 764–766 (2000).
3. Geeves, M. A. & Holmes, K. C. *Annu. Rev. Biochem.* **68**, 687–728 (1999).
4. Tanaka, H. *et al.* *Nature* **415**, 192–195 (2002).
5. Nishikawa, S. *et al.* *Biochem. Biophys. Res. Commun.* (in the press).
6. Rock, R. S. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 13655–13659 (2001).
7. Viegand, C., Wang, F., Bartoo, M. L., Sellers, J. R. & Molloy, J. E. *Nature Cell Biol.* **10**, 1038/ncb732 (2001).
8. Moore, J. R. *et al.* *J. Cell Biol.* **155**, 625–635 (2001).
9. Uyeda, T. Q. *et al.* *Proc. Natl Acad. Sci. USA* **93**, 4459–4464 (1996).
10. Anson, M. *et al.* *EMBO J.* **15**, 6069–6074 (1996).
11. Ruff, C., Furch, M., Brenner, B., Manstein, D. J. & Meyhöfer, E. *Nature Struct. Biol.* **8**, 226–229 (2001).
12. Walker, M. L. *et al.* *Nature* **405**, 804–807 (2000).
13. Wells, A. L. *et al.* *Nature* **401**, 505–508 (1999).

Global change

Getting cool with nitrogen

Allan H. Devol

Variations in the marine nitrogen cycle are implicated in driving glacial–interglacial climate change and producing warm spells during glacial periods. But phosphorus may also need to be taken into account.

Between the last glacial maximum, around 22,000 years ago, and the current interglacial period, the CO₂ content of the atmosphere rose by about 85 parts per million, or about 40%. The implication is that there was a shift of carbon from the deep ocean to the atmosphere, with the atmospheric carbon — in the form of CO₂ — exerting a greenhouse warming effect. How this carbon shift happened is not known, but it could have occurred through changes in phytoplankton photosynthesis in the oceans. Photosynthesis draws CO₂ out of the atmosphere and converts the carbon into organic form — when these organic molecules sink, they decompose and their carbon is locked up in the deep sea. So, in principle, reduced marine photosynthesis means that more CO₂ remains in the atmosphere and temperatures are higher.

In today's ocean, 'fixed' nitrogen (most of it in the form of NO₃⁻) and phosphorus (as PO₄³⁻) are required for phytoplankton photosynthesis in the ratio of 16:1, the Redfield ratio¹. Both nutrients are present at low concentrations that could limit photosynthesis. Nitrogen fixation is the process that converts gaseous N₂, which cannot be used by most marine plants, into readily available forms, and it has been suggested^{2,3} that this process has declined between the

last glacial maximum and now. The result would be less fixed nitrogen in the oceans and therefore, presumably, reduced photosynthesis, so potentially causing the deglacial rise in CO₂.

Two papers in this issue^{4,5} suggest alternative explanations. On page 156, Ganeshram *et al.*⁴ propose that changes in both the marine nitrogen and phosphorus cycles conspired to produce an N:P ratio that was greater than the Redfield ratio. This, they argue, would have suppressed nitrogen fixation and left the influence of photosynthesis on atmospheric CO₂ to be driven by small, slow increases in phosphorus concentration. On page 159, Altabet *et al.*⁵ present evidence of the importance of denitrification, the process that converts NO₃⁻ back to N₂ when oxygen is absent. They conclude that denitrification rates in the water column can change rapidly, altering the NO₃⁻ content of the ocean and thereby contributing to variations in atmospheric CO₂ levels.

Both groups^{4,5} use nitrogen-isotope data from sediment cores taken, respectively, from the ocean off western Mexico and from the Arabian Sea. These are two 'hot spot' areas in which denitrification zones are present over continental margins, and where denitrification preferentially removes the light isotope of nitrogen (¹⁴N) and enriches

the remaining NO_3^- in the heavy isotope (^{15}N). When this water is brought to the surface and the nutrients it contains are used in photosynthesis, the phytoplankton become enriched in the heavy isotope. This signal is then transferred to and preserved in sediments when the phytoplankton remains decompose. In both areas there are strong correlations between nitrogen-isotope ratio and estimates of the temperature record, with enriched values during warm periods and depleted values during cold periods.

The record presented by Altabet *et al.*⁵ extends back 65,000 years into the most recent ice ages. It is exceptionally detailed and the pattern it shows is strikingly similar to that of Dansgaard-Oeschger (D-O) events observed in Greenland ice cores⁶. These D-O events are warm periods that punctuated the ice ages roughly once every 3,000 years. Altabet *et al.* assume that the isotopic signals and D-O events are synchronous, and interpret the record as showing that, in a matter of centuries, conditions in the Arabian Sea have switched from there being little or no denitrification during cold periods to denitrification as intense as that which occurs today in warm periods. According to the alternative hypothesis, however, increased nitrogen fixation during the most recent glaciation would also have increased the marine nitrogen inventory and altered the isotopic balance, as observed, because nitrogen fixation adds fixed nitrogen depleted in the light isotope relative to the oceanic pool.

One thing that has perplexed palaeo-oceanographers is the asynchrony of climate change in the Antarctic and in Greenland⁷. Greenland temperature records during the most recent ice ages show 24 D-O warm periods, which are characterized by rapid warming and subsequent slow cooling⁶. Antarctic temperature variations show fewer, less pronounced warmings, and are rather gradual⁷. Altabet *et al.* suggest a possible mechanism for this. If the changes in denitrification inferred from their Arabian Sea cores indeed affected the nitrate inventory of the oceans (and, in turn, the CO_2 content of the atmosphere), then the time-scale of the inventory change would be damped by the residence time of nitrate in the ocean, which is currently estimated at about 3,000 years^{3,8}.

Smoothing their isotope record with a 3,000-year moving average, Altabet *et al.* have produced an isotope time series which, they argue, should reflect the marine nitrate inventory as being affected by denitrification in the waters of the Arabian Sea (which accounts for about one-third of the water-column denitrification in today's ocean). Between 65,000 and 25,000 years ago, this record is strikingly similar to Antarctic ice-core temperature and CO_2 records, implying that denitrification-driven variation in the

nitrogen cycle contributed to changing atmospheric CO_2 levels during this period.

One caveat, which is acknowledged by Altabet *et al.*, is that plankton must have been able to use this greater resource of fixed nitrogen during the most recent glaciation without a commensurate change in the phosphorus inventory—that is, the Redfield ratio was not fixed at 16:1. The hypothesis put forward by Altabet *et al.* is appealing. But because they have assumed that changes in denitrification were synchronous with D-O cycles, we cannot yet establish the timing of the changes in the nitrogen inventory, and thus determine whether they preceded or followed changes in atmospheric CO_2 and temperature. In other words, it cannot be said which was the cause of the other.

So has the glacial-interglacial increase in atmospheric CO_2 levels ultimately been driven by changes in the nitrogen cycle? It may depend on how rigidly Mother Nature enforces the Redfield ratio. If it is strictly enforced, as Ganeshram *et al.*⁴ suggest, then any increase in nitrate concentration in the oceans caused by reduced denitrification, or increased nitrogen fixation for that matter, would not affect marine photosynthesis without a commensurate change in phosphorus concentration. If, on the other hand, enforcement of the ratio is somewhat lax, then changes in the fixed nitrogen inventory, no matter how they are driven, and the consequential changes in marine carbon fixation, might well explain fluctuations in atmospheric CO_2 levels.

The place to look for the answers may again be the marine sedimentary record. The oceans are currently about 5‰ enriched in ^{15}N relative to atmospheric N_2 . Denitrification removes nitrate with an isotopic composition that is about 20‰ depleted in ^{15}N , whereas nitrogen fixation adds fixed nitrogen with a composition that is almost identical to that of atmospheric N_2 . So changes in the nitrate inventory that are driven by changes in either process should modify the ocean-wide isotopic composition of fixed nitrate and show up in certain places in the sedimentary record. Coaxing information on changes in the N:P ratio from this record, however, will be a more difficult challenge. ■

Allan H. Devol is in the School of Oceanography, University of Washington, Seattle, Washington 98195-7940, USA.

e-mail: devol@u.washington.edu

1. Redfield, A. C., Ketchum, B. H. & Richards, F. A. in *The Sea* Vol. 2 (ed. Hill, M. N.) 26–77 (Wiley, New York, 1963).
2. Falkowski, P. G. *Nature* **387**, 272–275 (1997).
3. Broecker, W. S. & Henderson, G. M. *Paleoceanography* **13**, 352–364 (1998).
4. Ganeshram, R. S., Pedersen, T. F., Calvert, S. E. & François, R. *Nature* **415**, 156–159 (2002).
5. Altabet, M. A., Higginson, M. J. & Murray, D. W. *Nature* **415**, 159–162 (2002).
6. Dansgaard, W. *et al.* *Nature* **364**, 218–220 (1993).
7. Blunier, T. *et al.* *Nature* **394**, 739–743 (1998).
8. Codispoti, L. A. *Nature* **376**, 724–725 (1995).

Daedalus

Hold on to your heat

Human skin loses a lot of heat. Clothes and fire are some of the most important inventions in our history, but they are still far from perfect. The 'goose-flesh' effect, in which a sudden frightening or frigid circumstance erects the local muscles and makes our skin hair stand on end, traps some insulating air but is feeble compared to the same process in cats, for example. We have to add or remove clothes several times a day, and often have difficulty getting to sleep because bedclothes have fixed heat loss.

So Daedalus is inventing a fabric with a variable insulation capacity. It has a current-carrying wire threading it, and magnetic fibrils wrapped around the wire. When a current flows, all the fibrils become magnetic the same way. They all repel each other, and stand up at right-angles to the wire. When the current ceases, they lie down again.

To minimize problems of power drain, the first product will be an eiderdown for a bed, accessible to the main power supply. Sleeping, that primitive need, often conflicts with the civilized need to stay up late. Furthermore, inflexible inventions such as the futon have made it harder to go to sleep. So our animal needs invite technical help. The DREADCO team's eiderdown has insulation that varies in different places and at different times. When the program is perfected, its user will get to sleep rapidly and stay asleep for longer. The resulting programmed bed might even be arranged to wake the sleeper up suddenly, by cooling at the right time. There may be no need for an alarm clock.

Working on the bed design will tell the team which fabric works best, what minimum current will increase insulation by a given factor, and so on. They will then be able to design simple dresses or trousers carrying their own rechargeable batteries. These should allow a wearer to move between rooms held at differing temperatures, or even to pop outside to the shops for a moment, without having to struggle into a coat.

But variable heat-loss garments pose a serious problem for all fashion designers. Underclothes are perhaps the easiest challenge; they allow considerable freedom for providing insulation. But the fashionable young must create the impression that they have endless body heat to spare. The design and placing of insulation panels in fashionable garments for the young will tax even DREADCO's designers to the limit.

David Jones

Predator and prey views of spider camouflage

Both hunter and hunted fail to notice crab-spiders blending with coloured petals.

Crab-spiders (*Thomisus onustus*) positioned for hunting on flowers disguise themselves by assuming the same colour as the flower, a strategy that is assumed to fool both bird predators and insect prey. But although this mimicry is obvious to the human observer, it has never been examined with respect to different visual systems. Here we show that when female crab-spiders mimic different flower species, they are simultaneously cryptic in the colour-vision systems of both bird predators and hymenopteran prey.

In animal communication, colouring is a compromise between being conspicuous to conspecifics and being poorly visible to predators or prey¹. Female crab-spiders adapt their entire body colour to that of the flowers on which they are trying to hide, a behaviour that is presumed to conceal them from predators and from the visiting pollinators that constitute their main prey².

To appear cryptic to both predators and prey, these spiders must precisely match the flower colour in their respective ranges of colour vision: four cone types, corresponding to ultraviolet (UV)–blue–green–red, for birds, and three (UV–blue–green) for insects^{3,4}. However, these visual systems differ markedly in their range of sensitivity and number of photoreceptors, making a precise colour match for both unlikely.

We used spectroradiometry to measure the degree of matching (crypsis) of *T. onustus* collected on the corollae of two flower species (*Senecio jacobaea* and *Mentha spicata*) growing around Tours, France. To analyse spider crypsis with respect to potential predators and prey on each flower, we computed and compared colour distances in the respective colour spaces of birds and Hymenoptera (see supplementary information).

For birds, spiders matched pink *Mentha* corollae when viewed through the four-cone colour-vision system (Fig. 1; $\chi^2=9$, d.f.=1, $P<0.01$). Likewise, on *Senecio*, each spider matched the individual yellow corolla on which it was waiting for prey ($\chi^2=7.5$, d.f.=1, $P<0.01$). But on the *Senecio* UV–yellow petal background, spiders produced a strong colour contrast that birds were likely to detect ($\chi^2=2.7$, d.f.=1, $P>0.05$). For Hymenoptera, spiders matched the blue-green colour of *Mentha* corollae in the three-cone colour-vision system ($\chi^2=4$, d.f.=1, $P<0.05$). They also effectively mimicked the individual blue-green colour of *Senecio* corollae ($\chi^2=7.5$, d.f.=1, $P<0.01$), but contrasted on UV–green petals when seen by Hymenoptera

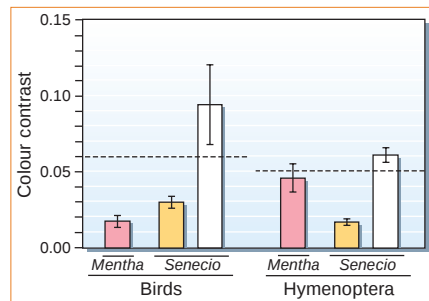


Figure 1 Colour contrast of spiders (mean euclidean distances $\pm$ s.e.m.) against *Mentha* corollae (pink bars), *Senecio* corollae (yellow bars) and *Senecio* petals (white bars) when viewed by birds and Hymenoptera. Dashed lines indicate thresholds for colour-contrast detection calculated for birds and Hymenoptera. For details of modelling calculations, see supplementary information.

($\chi^2=2.7$, d.f.=1, $P>0.05$).

To detect small targets, birds and bees can use achromatic vision instead of colour contrast^{5,6}. Spiders were significantly brighter than corollae of *Mentha* (repeated-measures ANOVA, birds: $F_{1,17}=8.5$, $P<0.001$; Hymenoptera: $F_{1,17}=8.6$, $P<0.001$) and *Senecio* (birds: $F_{1,14}=35.7$, $P<0.001$; Hymenoptera: $F_{1,14}=46.2$, $P<0.001$), but were darker than *Senecio* petals (birds: $F_{1,14}=133.7$, $P<0.001$; Hymenoptera: $F_{1,14}=157.3$, $P<0.001$).

Our results indicate that crab-spiders' colour mimicry works successfully on the visual systems of both predator and prey, achromatic vision being potentially more efficient under particular viewing conditions. This aggressive mimicry may vary from species to species, as shown by *Misumena vatia*⁷, a crab-spider that reduces its chromatic contrast to bees on white flowers (as in Fig. 2), as does *T. onustus*, but



Figure 2 Master of disguise: a crab-spider (left), concealed against a flower's white petals, preys on an unsuspecting bee.

is also able to reduce its achromatic contrast on yellow flowers.

Marc Théry*, Jérôme Casas†

*Laboratoire d'Ecologie Générale, Muséum National d'Histoire Naturelle, Centre National de la Recherche Scientifique UMR 8571, 91800 Brunoy, France
e-mail: thery@mnhn.fr

†Institut de Recherche sur la Biologie de l'Insecte, Université de Tours, Centre National de la Recherche Scientifique UMR 6035, 37200 Tours, France

1. Bradbury, J. W. & Vehrencamp, S. L. *Principles of Animal Communication* (Sinauer, Sunderland, Massachusetts, 1998).
2. Oxford, G. S. & Gillespie, R. G. *Annu. Rev. Entomol.* **43**, 619–643 (1998).
3. Chen, D.-M. & Goldsmith, T. H. *J. Comp. Physiol. A* **159**, 473–479 (1986).
4. Briscoe, A. D. & Chittka, L. *Annu. Rev. Entomol.* **46**, 471–510 (2001).
5. Osorio, D., Miklósi, A. & Gonda, Z. *Evol. Ecol.* **13**, 673–689 (1999).
6. Spaethe, J., Tautz, J. & Chittka, L. *Proc. Natl Acad. Sci. USA* **98**, 3898–3903 (2001).
7. Chittka, L. *Entomol. Gener.* **25**, 181–187 (2001).

Supplementary information accompanies this communication on Nature's website (www.nature.com).

Competing financial interests: declared none.

COMMUNICATIONS ARISING Brain evolution

Analysis of mammalian brain architecture

The mammalian brain is composed of several distinct parts which show different growth in evolution. Clark, Mitra and Wang¹ found that the two main cortices of the brain — the cerebral (neo-) cortex and the cerebellum — show very different growth, and that whereas the ratio of neocortex volume to total brain volume increases with evolution, the cerebellum occupies a constant proportion in different species. Here I compare the surface areas of

the two cortices in different species and find that these show a simple proportionality. Contrary to the conclusion drawn by Clark *et al.*¹, this linear dependence of size implies that the two major cortices increase their computational capacity in parallel, suggesting a functional dependence of the one upon the other.

The results of Clark *et al.*¹ are unexpected, because it is known that the cerebellar parts that show the most pronounced growth in evolution are linked to the neo-cortex and that, congruently, the main input and output structures of the cerebellum (the pontine and dentate nuclei) are also closely linked to the neocortex in evolutionary growth². Together with the

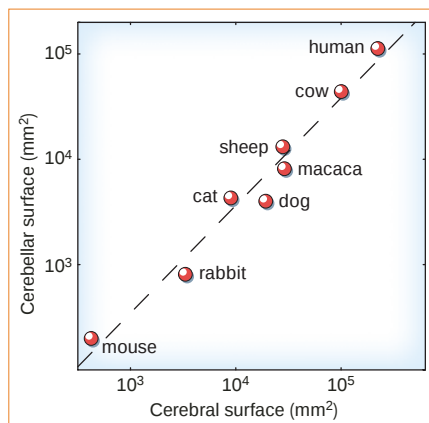


Figure 1 Logarithmic plot of cerebellar¹¹ versus cerebral cortex area¹². The slope of the least-squares logarithmic regression line is 1.02 ($y = 1.02x - 0.52$, Spearman's $r = 0.95$, $P < 0.0003$), indicating a linear dependence of cerebellar on cerebral cortex surface. (Brodmann's data provided by H. J. Jerison.)

neocortex, these two nuclei are among the brain structures that show the largest increase in primates². This interconnect- edness between the two cortices does not appear in the measurements of Clark and colleagues¹.

I have considered cortical surface areas, rather than volumes, as a potentially more appropriate measure of the computational capacity of sheet-like formations, and observe a linear relation between telen- cephalon and cerebellum over a wide range of different mammals (Fig. 1). As the cerebellum is a highly folded cortex of almost uniform thickness in all species, the increase in cerebellar size can also be measured in terms of cortical surface area.

However, the cerebral cortex increases in thickness by a factor of 7.6 across different mammalian species³, making the increase in volume in larger species greater than proportional to the increase in area. This should be taken into account when consid- ering the relative volumes of the two cortices. In the case of the cerebellum, its volume represents a constant fraction of the total brain volume, as shown by Clark *et al.*¹.

The discrepancy between the results of Clark *et al.*¹ and the constant ratio of cerebral and cerebellar surface areas may also be due to their inclusion of the white matter in the volumes of neocortex and cerebellum. The human neocortex consists of nearly 42% white matter, a peak value for primates, in which, as in other mammals, larger brains have a greater proportion of white matter⁴. In contrast, the cerebellum has a more-or- less constant proportion of white matter (30% in rats⁵ and 26% in humans⁶).

The increase in neocortical white matter during evolution^{7,8} probably reflects the need in larger brains to maintain the high connectivity required to operate associative networks⁹. The cerebellum, in contrast, lacks a system of intrinsic long-range connections and relies entirely on the opera-

tion of local short-range connections within the grey matter¹⁰. The cerebellar subcortical white matter is composed only of input and output fibres, the number of which is pro- portional to the surface area of the cortex.

Fahad Sultan

Department of Cognitive Neurology,
University of Tübingen, Auf der Morgenstelle 15,
72076 Tübingen, Germany
e-mail: fahad.sultan@uni-tuebingen.de

1. Clark, D. A., Mitra, P. P. & Wang, S. S. *Nature* **411**, 189–193 (2001).
2. Matano, S., Stephan, H. & Baron, G. *Folia Primatol.* **44**, 171–181 (1985).
3. Hofman, M. A. *Brain Behav. Evol.* **27**, 28–40 (1985).
4. Frahm, H. D., Stephan, H. & Stephan, M. J. *Hirnforsch.* **23**, 375–389 (1982).
5. Korbo, L., Andersen, B. B., Ladefoged, O. & Møller, A. *Brain Res.* **609**, 262–268 (1993).
6. Andersen, B. B., Korbo, L. & Pakkenberg, B. J. *Comp. Neurol.* **326**, 549–560 (1992).
7. Braitenberg, V. J. *Comput. Neurosci.* **10**, 71–77 (2001).
8. Zhang, K. & Sejnowski, T. J. *Proc. Natl Acad. Sci. USA* **97**, 5621–5626 (2000).
9. Braitenberg, V. & Schüz, A. *Anatomy of the Cortex* (Springer, Berlin, 1991).
10. Braitenberg, V., Heck, D. & Sultan, F. *Behav. Brain Sci.* **20**, 229–245 (1997).
11. Sultan, F. & Braitenberg, V. J. *Hirnforsch.* **34**, 79–92 (1993).
12. Brodmann, K. *Verhandlungen der 85. Versammlung deutscher Naturforscher und Ärzte in Wien* **85**, 200–240 (1913).

How did brains evolve?

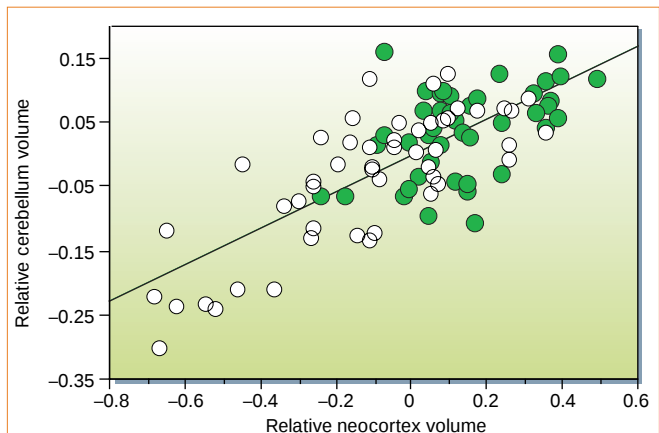
Three reports on mammalian brain evolu- tion^{1–3} analyse the same comparative data on brain component volumes⁴ but come to partially conflicting conclusions. Clark *et al.*³ conclude from their analysis of volumetric brain proportions (“cerebro- types”) that cerebellum size is invariant across mammalian taxonomic groups, the neocortex and cerebellum do not co-vary in size (in contradiction to ref. 1), and cerebrotype-based measures identify direc- tional changes in brain architecture. Here I provide evidence that calls each of these conclusions into question. The failure of the cerebrotype measure to identify species

differences in brain architecture that are independent of gross brain size undermines the proposal by Clark *et al.* that it could be useful for detecting evolutionary patterns and phylogenetic relationships.

In attempting to establish uniformity of cerebellum size, Clark *et al.* do not use a multiple-comparisons procedure, instead carrying out *t*-tests for each taxon against all the other taxa, thereby pooling taxa with smaller- and larger-than-average cerebellum size. In contrast, analysis of variance (ANOVA) on cerebellar volume proportion in nine mammalian orders⁵ (not including two echolocating taxa said to deviate from cerebellar constancy³) indicates significant variance ($F = 18.8$, d.f. = 7, 81, $P < 0.0001$, 15 of 28 pairwise comparisons significant). What Clark *et al.* have observed is that neocortex size varies more than cerebellum size, so that variation in the latter is small as a proportion of the whole. This does not, however, contradict an important evolu- tionary relationship between them.

The fact that, as the neocortical fraction of brain size increases, the cerebellum does not, like all the other structures, decrease as a proportion of total brain volume³, suggests that the cerebellum and neocortex evolved together, but with the cerebellum evolving more slowly. When variation in the size of other brain structures is partial- led out, there is a significant correlation between cerebellum and neocortex size (Fig. 2). This correlation is not dependent on using residuals from linear regression, as it is also found using simple ratios of cerebellum and neocortex size to the size of the rest of the brain (across taxa, $r^2 = 0.87$, d.f. = 1, 91, $P < 0.0001$). Neither is it an artefact of taxonomic effects such as ‘grade shifts’¹, as it is apparent using the method of phylogenetically independent contrasts⁶, which controls for such effects (primates, $r^2 = 0.33$, d.f. = 1,39, $P < 0.0001$; insecti- vores, $r^2 = 0.27$, d.f. = 1,32, $P = 0.002$). Data

Figure 2 Correlated variation in the relative size of neocortex and cerebellum. Relative size was defined as the residual from the least-squares regression of structure volume on the volume of the rest of the brain. The correlation is significant across taxa ($r^2 = 0.57$ d.f. = 1,91, $P < 0.0001$, slope = 0.29), and also within each taxon (pri- mates, filled circles, $r^2 = 0.22$, $P = 0.001$, slope = 0.20; insecti- vores, hollow circles, $r^2 = 0.62$, $P < 0.001$, slope = 0.24). The correlations are strengthened when the diencephalon, through which neocortex–cerebellum connections project, is excluded from the rest of the brain (for example, across taxa, $r^2 = 0.72$, $P < 0.0001$, slope = 0.35). In each case, slopes of substantially less than unity indicate that cerebellum size evolved more slowly than neocortex size.



on bats⁷ also indicate a positive correlation between residual neocortex and cerebellum size ($n = 158$ species, $r^2 = 0.45$, $P < 0.0001$). Hence different analyses in several taxa all point to a relationship between relative neocortex and cerebellum size.

A problem with using volume proportions³ is that these correlate with overall brain size in all taxa studied, and hence conflate allometric trends (nonlinear scaling) among brain components⁸ with variation in component size that is independent of whole-brain variation¹. Clark *et al.* mistakenly claim that volume fractions among insectivores do “not vary systematically with brain size for any principal developmental brain division”³. As in primates, neocortex proportion is positively correlated with brain size among insectivores (linear regression on log-transformed brain volume, $n = 50$ species; $r^2 = 0.18$, $P = 0.002$) and also in bats⁷ ($r^2 = 0.58$, $n = 158$ species, $P < 0.0001$), suggesting that this is not an unusual feature of primates associated with “directed selection pressure” in that taxon³.

This conflation of brain size and size-independent structural differences limits the utility of the cerebrotypic measure for evaluating adaptive patterns and phylogenetic relationships. For example, cerebrotypes place gibbons (*Hylobates*) with similarly sized Old World monkeys and not with their closest phylogenetic relatives, the other apes³.

Clark *et al.* suggest that selection has caused important shifts in brain proportions at the origin of major mammalian taxa, whereas constraints have resulted in “a relatively uniform cerebrotypic” within each taxon³. In contrast, another study concludes that mosaic brain organization is caused by selective adaptation within orders, whereas between orders there is an interplay between selection and constraints². However, it is unnecessary to postulate that constraints have been more active either at the origin of major taxa or during their subsequent evolutionary radiation. Clark *et al.* interpret the partly non-overlapping distributions of their multivariate representations of brain proportions as evidence that shifts in cerebrotypic evolved at infrequent (about 10-million-year) intervals, followed by relative stasis. However, these distributions would be expected from an evolutionary pattern of gradual divergence and extinction of some intermediate forms. The most parsimonious assumption is that both selection and constraints operated similarly at different taxonomic levels.

Robert A. Barton

Evolutionary Anthropology Research Group,
Department of Anthropology, University of
Durham, Durham DH1 3HN, UK
e-mail: r.a.barton@durham.ac.uk

- 189–193 (2001).
4. Stephan, H., Frahm, H. D. & Baron, G. *Folia Primatol.* **35**, 1–29 (1981).
5. Putnam, I. K. *Proc. R. Acad. Amsterdam* **31**, 155–168 (1927).
6. Purvis, A. & Rambaut, A. *Comp. Appl. Biosci.* **11**, 247–251 (1995).
7. Baron, G., Stephan, H. & Frahm, H. D. *Comparative Neurobiology in Chiroptera* Vol. 1 (Birkhauser, Basel, 1996).
8. Finlay, B. L. & Darlington, R. B. *Science* **268**, 1578–1584 (1995).

Clarke *et al.* reply — Sultan's observations do not contradict our finding that, when normalized to the whole brain, cerebellar volume is relatively constant. To make interspecies comparisons, we used the volume fraction (the volume of a component divided by the total brain volume)¹. Without this normalization, the dominant trend is variation in absolute size². This logic applies to surface areas as well, and Sultan's interesting analysis reflects this principle.

Sultan points out that because cerebellum and neocortex differ fundamentally in their basic wiring structure, neocortical white matter should be excluded from our calculations. Doing this increases the effective cerebellar (cbl) volume fraction to $F_{cbl}' = F_{cbl} / (1 - F_w)$, where F_w is the neocortical white-matter volume fraction. We made this correction in 44 species^{3,4} and found that its effect was negligible: F_{cbl}' is still relatively constant (0.16 ± 0.04 , mean $\pm$ s.d., 13 taxa), whereas the corrected neocortical grey-matter volume fraction is not. The quantities are uncorrelated ($r^2 = 0.07$, d.f. = 11, $P = 0.4$). The cerebellar correction is largest in cetaceans ($F_{cbl}' = 0.25 \pm 0.06$, 2 species), emphasizing our finding that echolocating mammals have unusually large cerebellar volume fractions. Note that the use of a *t*-test (not ANOVA) was appropriate here because we wanted to know whether a taxon's mean was outlying.

Barton's allometric approach to quantifying size differences relies on fitting data to a single power law (a linear fit in log-log coordinates). He uses data on neocortical volume (N), cerebellar volume (C), and the rest of the brain ($B - N - C$). He fits for $\log N$ and $\log C$ against $\log (B - N - C)$ by pooling primate and insectivore taxa; the fit slope is 1.62 for neocortex and 1.28 for cerebellum. But subtracting these fits discards the difference between these two baseline relationships, which constitutes the dominant size trend between neocortex and cerebellum. As the underlying reason for the fitted trend is unknown, it is unclear whether the resulting residuals are interpretable.

Calculating residuals creates another problem: pooled data from multiple taxa are generally not well fitted by a single power law^{1,5}. As a result, residuals calculated to a single line contain systematic errors (see Fig. 13 of ref. 1 and Fig. 4 of ref. 5). For instance, fits for different taxa often have similar slopes but different intercepts, a

phenomenon known as grade shifting. This is apparent in Barton's analysis, in which nearly all primates fall above the fit lines, causing neocortical and cerebellar residuals to be mostly positive for primates and negative for insectivores (Fig. 2), and leading to an artefactual correlation.

To remedy this, we considered the primates separately, dividing them into lemurs and lorises (haplorhines) and tarsiers, monkeys and apes (strepsirhines). After fitting each group to its own power law, we recalculated neocortical and cerebellar residuals and found that one set of residuals can explain only 9% of the variance in the other set ($r^2 = 0.09$, 39 d.f., $P = 0.1$). Recalculation including insectivores gives similar results, although classification of insectivores poses more difficulties than primates^{6,7}. Therefore, the correlation reported by Barton does not exist within taxa; it largely (although indirectly) reflects differences in brain architecture among taxa. The same arguments also apply to his other analyses.

A third type of error arises when components are combined for analysis. The sum of two power laws of differing exponent can never yield a power law. Therefore any analysis (including Barton's) that combines multiple components (for instance, the rest of the brain, or neocortex equals grey plus white matter) intrinsically contradicts the power-law assumption⁸.

Underlying our disagreements with Barton is the fact that allometric relationships constitute *ad hoc* models that often provide only a rough fit to the data⁵. Where power-law phenomena have an explanation, such as in the circulatory system, the tissue in question is substantially more homogeneous than the whole brain. For this reason, using residuals based on subtracting approximate power-law trends may not be particularly effective at identifying relationships in the architecture of mammalian brains.

Samuel S.-H. Wang*, Partha P. Mitra†, Damon A. Clark*‡

Departments of *Molecular Biology and †Physics,
Princeton University, Princeton,
New Jersey 08544, USA

e-mail: samwang@princeton.edu

‡Bell Laboratories, Lucent Technologies,
Murray Hill, New Jersey 07974, USA

1. Jerison, H. J. *Brain Size and the Evolution of Mind* (Am. Mus. Natl Hist., 1991).
2. Finlay, B., Darlington, R. & Nicastro, N. *Behav. Brain Sci.* **24**, 263–278 (2001).
3. Frahm, H. D., Stephan, H. & Stephan, M. J. *Hirnforsch.* **23**, 375–389 (1982).
4. Hofman, M. A. *Brain Behav. Evol.* **32**, 17–26 (1988).
5. Gould, S. J. in *Approaches to Primate Paleobiology* (ed. Szalay, F. S.) 244–292 (Karger, Basel, 1975).
6. Stanhope, M. J. *et al. Proc. Natl Acad. Sci. USA* **95**, 9967–9972 (1998).
7. Mouchaty, S. K., Gullberg, A., Janke, A. & Arnason, U. *Mol. Biol. Evol.* **17**, 60–67 (2000).
8. Zhang, K. & Sejnowski, T. J. *Proc. Natl Acad. Sci. USA* **97**, 5621–5626 (2000).

Altruistic punishment in humans

Ernst Fehr* & Simon Gächter†

* University of Zürich, Institute for Empirical Research in Economics, Blümlisalpstrasse 10, CH-8006 Zürich, Switzerland

† University of St Gallen, FEW-HSG, Varnbühlstrasse 14, CH-9000 St Gallen, Switzerland

Human cooperation is an evolutionary puzzle. Unlike other creatures, people frequently cooperate with genetically unrelated strangers, often in large groups, with people they will never meet again, and when reputation gains are small or absent. These patterns of cooperation cannot be explained by the nepotistic motives associated with the evolutionary theory of kin selection and the selfish motives associated with signalling theory or the theory of reciprocal altruism. Here we show experimentally that the altruistic punishment of defectors is a key motive for the explanation of cooperation. Altruistic punishment means that individuals punish, although the punishment is costly for them and yields no material gain. We show that cooperation flourishes if altruistic punishment is possible, and breaks down if it is ruled out. The evidence indicates that negative emotions towards defectors are the proximate mechanism behind altruistic punishment. These results suggest that future study of the evolution of human cooperation should include a strong focus on explaining altruistic punishment.

Throughout evolution, crucial human activities like hunting big game, sharing meat, conserving common property resources, and warfare constituted a public good. In situations like these, every member of the group benefits from the 'good', including those who did not pay any costs of providing the good. This raises the question of why people regularly participate in costly cooperative activities like warfare and big-game hunting^{1–4}. Several theories have been proposed to explain the evolution of human cooperation. The theory of kin selection⁵ focuses on cooperation among individuals that are genetically closely related, whereas theories of direct reciprocity^{6–9} focus on the selfish incentives for cooperation in bilateral long-term interactions. The theories of indirect reciprocity^{10–15} and costly signalling^{16–18} show how cooperation in larger groups can emerge when the cooperators can build a reputation. Yet these theories do not readily explain why cooperation is frequent among genetically unrelated people, in non-repeated interactions, when gains from reputation are small or absent.

Punishment provides a solution to this problem. If those who free ride on the cooperation of others are punished, cooperation may pay^{3,19–23}. Yet this 'solution' begs the question of who will bear the cost of punishing the free riders. Everybody in the group will be better off if free riding is deterred, but nobody has an incentive to punish the free riders. Thus, the punishment of free riders constitutes a second-order public good. The problem of second-order public goods can be solved if enough humans have a tendency for altruistic punishment, that is, if they are motivated to punish free riders even though it is costly and yields no material benefits for the punishers.

We examined the question of whether humans engage in altruistic punishment and how this inclination affects the ability of achieving and sustaining cooperation. A total of 240 students participated in a 'public goods' experiment with real monetary stakes and two treatment conditions: punishment and no punishment. In both conditions, groups with four members played the following public goods game. Each member received an endowment of 20 money units (MUs) and each one could contribute between 0 and 20 MUs to a group project. Subjects could keep the money that they did not contribute to the project. For every MU invested in the project, each of the four group members, that is, also those who invested little or nothing, earned 0.4 MUs. Thus, the investor's return from investing one additional MU in the project was 0.4 MUs, whereas the group return was 1.6 MUs. Because the cost of investing 1 MU in the project was exactly 1 MU, whereas the private return was only 0.4 MUs, it was always in the material self-interest of any subject to keep all MUs privately—irrespective of how much the other three subjects contributed. Yet, if all group members kept all MUs

privately, each subject earned only 20 MUs, whereas if all of them invested their 20 MUs each subject would earn $0.4 \times 80 = 32$ MUs.

All the interactions in the experiment took place anonymously. Members were not informed of the identity of the others in the group. Subjects made their investment decisions simultaneously and, once the decisions were made, they were informed about the investments of the other group members. The only difference between the two conditions was that in the punishment condition, subjects could punish each of the other group members after they were informed about the others' investments. A punishment decision was implemented by assigning between 0 and 10 points to the punished member. Each point assigned cost the punishing member 3 MUs and the punished member 1 MU. All the punishment decisions were also made simultaneously.

Because we conjectured that the opportunity for punishing would have a larger impact if subjects could learn about the behaviour of other group members, we repeated the basic public goods game—with and without punishment opportunity, depending on the treatment—for six periods. To rule out that repetition created cooperation or punishment through direct reciprocity^{6–9} or reputation^{10–15}, the group composition changed from period to period such that no subject ever met another subject more than once. Moreover, our design ruled out any kind of reputation formation (see Methods), so purely selfish subjects will never cooperate or punish others, because cooperation and punishment are costly and yield no pecuniary benefits. Therefore, the selfish motives associated with theories of indirect reciprocity^{10–15} or costly signalling^{16–18} cannot explain cooperation and punishment in this environment.

However, punishment may well benefit the future group members of a punished subject, if that subject responds to the punishment by raising investments in the following periods. In this sense, punishment is altruistic. In the presence of altruistic punishers, even purely selfish subjects have a reason to cooperate in the punishment treatment.

Altruistic punishment and cooperation

Altruistic punishment took place frequently. In the ten sessions, subjects punished other group members a total of 1,270 times; 84.3% of the subjects punished at least once, 34.3% punished more than five times during the six periods, and 9.3% punished even more than ten times. Punishment also followed a clear pattern. Most (74.2%) acts of punishment were imposed on defectors (that is, below-average contributors) and were executed by cooperators (that is, above-average contributors), and punishment of the defectors was harsh (Fig. 1). For example, if a subject invested

14–20 MUs less than the average investment of the other group members during periods 5 and 6, the total group expenditures for punishing this subject were almost 10 MUs. Moreover, the more a subject's investment fell short of the average investment of the other three group members, the more the subject was punished. The pattern and strength of punishment was also stable across time (Fig. 1). A Wilcoxon signed rank test of punishment in periods 1–4 versus periods 5 and 6, with 10 matched observations, yields $z = -1.07$, $P = 0.285$ (two-tailed). The same test for periods 1–5 versus period 6 yields $z = 0.178$, $P = 0.859$ (two-tailed).

We examined how the group expenditures for the punishment of member i varied with the positive and the negative deviation of member i 's cooperation from the average cooperation of the others. Our examination is based on Tobit regressions. The regression coefficient on 'negative deviation' is 0.622 ($z = 18.1$, $P < 0.0005$) and the coefficient on 'positive deviation' is -0.149 ($z = -2.86$, $P < 0.004$). The hypothesis tests associated with the regression are based on robust standard errors that take into account that only the observations across sessions are independent and that punishment is a censored variable. The average cooperation of other group members, if added as an explanatory variable to the regression, is insignificant ($z = -0.99$, $P = 0.322$). This analysis indicates that an increase in the negative deviation of i from the others' average cooperation by 10 MUs increased the punishment expenditure of the others by 6.22 MUs (and, hence, the pay-off reduction imposed on i by 18.66 MUs), whereas an increase in the positive deviation of i by 10 MUs reduced the punishment expenditure by 1.49 MUs. This punishment pattern led to a hump-shaped relation between an individual's income and the deviation from the average cooperation of the other group members. The income was highest when the individual's investment was close to the average investment of the others. Both positive and negative deviations from the average investment decreased an individual's income.

The punishment of non-cooperators substantially increased the amount that subjects invested in the public good. In the five sessions where the punishment condition was the first treatment (Fig. 2a), the average cooperation level was much higher in the punishment condition (Wilcoxon signed rank test, five matched observations, $z = -2.023$, $P = 0.043$, two-tailed). The average investment of 94.2% of the subjects was higher in the punishment condition. In fact, the average investment in the punishment condition was higher in each session and in each period than the average investment in

any of the periods and sessions of the no-punishment condition. The time trend of the average investment was also rather different in the two conditions (Fig. 2a). Although cooperation increased over time in the punishment condition, it sharply decreased in the no-punishment condition. In the final period of the punishment condition, 38.9% of the subjects contributed their whole endowment and 77.8% contributed 15 MUs or more. In the final period of the no-punishment condition, 58.9% of the subjects contributed nothing and 75.6% contributed 5 MUs or less.

A very similar pattern emerged in the sessions where the no-punishment condition came first (Fig. 2b). The average cooperation again was much higher in the punishment condition (Wilcoxon signed rank test, five matched observations, $z = 2.023$, $P = 0.043$, two-tailed). In the punishment condition, 91.4% of the subjects contributed more than in the no-punishment condition. In addition, although the average investment decreased in the no-punishment condition, it increased sharply in the punishment condition. Moreover, Fig. 2b indicates that the punishment threat was immediately effective because there was a large upwards jump in investments when the punishment opportunity was made available to the subjects. It also turns out that the sequence of the treatments had no effect on cooperation. Investments in the punishment condition were similar, irrespective of whether this condition came first or second in a session (Mann–Whitney test, $z = 0.104$, d.f. = 4, $P = 0.918$, two-tailed). The same held for the no-punishment condition (Mann–Whitney test, $z = 1.358$, d.f. = 4, $P = 0.175$, two-tailed). Thus, we can use the data of all ten sessions to compare the average investments across conditions so that the differences are significant at a much higher level (Wilcoxon signed rank test, ten matched observations, $z = 2.803$, $P = 0.005$, two-tailed).

It is not only the punishment opportunity (that is, the non-executed punishment threat) but also the actual punishment that raised cooperation levels. When a subject was punished before period 6, that subject raised investment in the next period on average by 1.62 MUs. Note, however, that this does not constitute an indirect material benefit of the act of punishment for an individual punisher, because the punishing subject never meets

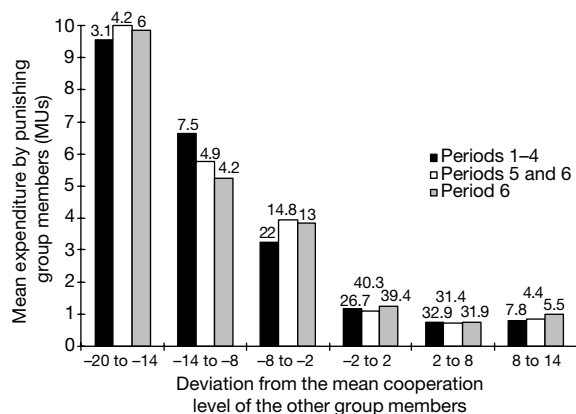


Figure 1 Mean expenditure on punishment during different time intervals as a function of the deviation of the cooperation of the punished group member from the mean cooperation of the other members. Each money unit (MU) spent on punishment reduced the income of the punished member by 3 MUs. The numbers above the bars indicate the relative frequency of the observations underlying the bars. For example, during periods 1–4, individual group members deviated between -20 and -14 MUs from the average cooperation of other group members in 3.1% of all cases.

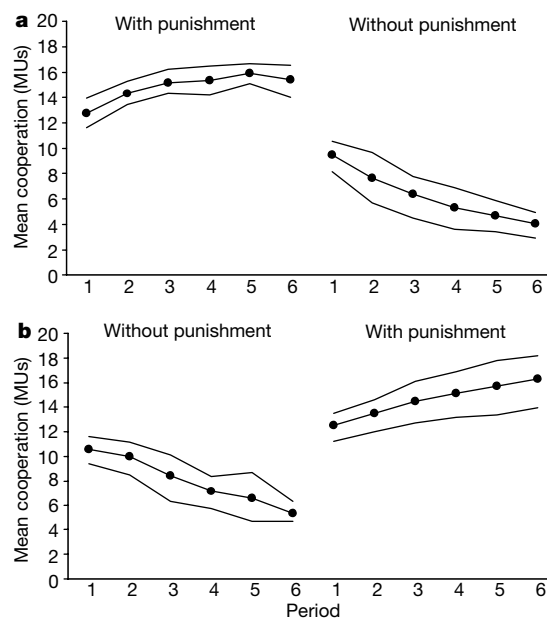


Figure 2 Time trend of mean cooperation together with the 95% confidence interval. **a**, During the first six periods, subjects have the opportunity to punish the other group members. Afterwards, the punishment opportunity is removed. **b**, During the first six periods, punishment of other group members is ruled out. Afterwards, punishment is possible.

the same subjects again. The act of punishment does provide a material benefit for the future interaction partners of the punished subject but not for the punisher. Thus, the act of punishment, although costly for the punisher, provides a benefit to other members of the population by inducing potential non-cooperators to increase their investments. For this reason, the act of punishment is an altruistic act.

Emotions as a proximate mechanism

Given the pattern of punishment, the investment behaviour of subjects seems quite rational. To avoid punishment, subjects invested in accordance with the group norm. But we wondered why subjects punish free riders in a one-shot context when this is costly. With regard to the proximate source of the punishment, negative emotions may provide an explanation. Free riding may cause strong negative emotions among the cooperators and these emotions, in turn, may trigger their willingness to punish the free riders^{24,25}. If this conjecture is correct, we should observe particular emotional patterns in response to free riding. To elicit these patterns, the participants were confronted with the following two hypothetical investment scenarios after the final period of the second treatment (the numbers in brackets relate to the second scenario):

"You decide to invest 16 [5] francs to the project. The second group member invests 14 [3] and the third 18 [7] francs. Suppose the fourth member invests 2 francs to the project. You now accidentally meet this member. Please indicate your feeling towards this person."

After they had read a scenario, subjects had to indicate the intensity of their anger and annoyance towards the fourth person (the free rider) on a seven-point scale (1 = 'not at all' to 7 = 'very much'). The difference between scenarios 1 and 2 is that the other three persons in the group contribute relatively much in scenario 1 and relatively little in scenario 2. It turns out that a free rider triggered much anger among the other subjects if these subjects contributed a lot relative to the free rider (scenario 1). Forty-seven per cent of the subjects indicated an anger level of 6 or 7 and another 37% indicated an anger level of 5. If the deviation of the free rider's contribution from the other members' contribution was relatively small (scenario 2), the anger level was significantly lower (Wilcoxon signed rank test, $z = 9.636$, $P < 0.0005$) but still considerable. In this case (scenario 2), 17.4% of the subjects indicated an anger level of 6 or 7 and 80.5% indicated a level of 4 or 5 in scenario 2. This shows that the intensity of negative emotions towards a free rider varies with the deviation from the others' average contribution.

Because we were also interested in the free riders' expectation of the other members' anger, we confronted subjects with a third and a fourth hypothetical scenario (numbers in brackets relate to scenario 4):

"Imagine that the other three group members invest 14, 16 and 18 [3, 5 and 7] francs to the project. You invest 2 francs to the project and the others know this. You now accidentally meet one of the other members. Please indicate the feelings you expect from this member towards you."

In scenarios 3 and 4, the hypothetical free rider had to indicate the expected anger of the others on a seven-point scale. The anger that was expected by the free riders in scenario 3 was even greater than the actually expressed anger according to scenario 1 (Wilcoxon signed rank test, $z = 7.68$, $P < 0.0005$). In scenario 3, 74.5% of the subjects expected the anger level of others to be 6 or 7, and 22.5% expected an anger level of 5. In scenario 4, the deviation of the hypothetical free rider from the others' contribution was smaller than in scenario 3. This decrease in the deviation from the others caused significant differences in expected anger levels between scenarios 3 and 4 (Wilcoxon signed rank test, $z = 12.17$, $P < 0.0005$). Only 17.8% of the hypothetical free riders expected anger levels of 6 or 7 in scenario 4 and 80% expected levels of 4 or 5.

The low contributors in the no-punishment condition expected a higher intensity of negative emotions than the high contributors. This probably reflects the fact that the low contributors in the no-punishment condition experienced more sanctions in the punishment condition.

The same four scenarios were presented to 33 subjects that had not previously participated in our experiments, to check whether participating in the experiment affects the elicited emotions. The same emotional patterns that were expressed by our 240 experimental subjects were expressed by the 33 subjects who did not participate in our games.

Our results suggest that free riding causes strong negative emotions and that most people expect these emotions. Moreover, the above emotional pattern is consistent with the hypothesis that emotions trigger punishment for the following reasons. First, if negative emotions trigger punishment, most punishment acts would be expected to be executed by above-average contributors and imposed on below-average contributors. This is clearly the case in our experiments: 74.2% of all punishment acts follow this pattern. Second, punishment increased with the deviation of the free rider from the average investment of the other members. This is exactly what would be expected if negative emotions are the proximate cause of the punishment, because negative emotions became more intense as the free rider deviated further from the others' average investment. Third, if negative emotions cause punishment, the punishment threat is rendered immediately credible because most people are well aware that they trigger strong negative emotions when they free ride. Therefore, we should detect an immediate impact of the punishment opportunity on contributions at the switch points between the punishment and the no-punishment condition. This is what we observed. The introduction (or elimination) of the punishment opportunity led to an immediate rise (or fall) in investment (see Fig. 2). Taken together, these observations are consistent with the view that emotions are an important proximate factor behind altruistic punishment.

Our evidence has profound implications for the evolutionary study of human behaviour. In the past, human cooperation has mainly been explained in terms of kin selection, reciprocal altruism, indirect reciprocity and costly signalling. These theories focus attention on mechanisms other than altruistic punishment. By showing that altruistic punishment is a key force in the establishment of human cooperation, our study indicates that there is more at work in sustaining human cooperation than is suggested by these theories. Thus, our evidence suggests that the evolutionary study of human cooperation in large groups of unrelated individuals should include a focus on explaining altruistic punishment^{3,26,27}. Moreover, because altruistic punishment occurs among genetically unrelated individuals and under conditions that rule out direct reciprocity and reputation formation, the above-mentioned theories do not readily account for altruistic punishment. □

Methods

A total of 240 undergraduate students (31% females) from the University of Zürich and the Federal Institute of Technology (ETH) voluntarily participated in the experiments. Special care was exerted to recruit students from many different disciplines to maximize the chances that the subjects had never met before. Ten experimental sessions with 24 subjects took place. Each of the 24 subjects played two 6-period public goods games: a game without a punishment opportunity and a game with a punishment opportunity. In five sessions, subjects first played the punishment treatment and then the no-punishment treatment; in the other five sessions, the treatment sequence was reversed. When subjects played the first six-period game, they did not know that another game would take place after period 6. At the beginning they were informed that the experiment would last for six periods. After period 6, subjects were told that another six-period experiment would take place and that thereafter the whole session would be over. The experiments typically lasted 60 min and on average subjects earned 39.7 Swiss francs (US\$23.95, 27.2 euros) per session.

In each period of a session, the 24 subjects were randomly allocated to six groups of four subjects. The allocation of subjects to the groups ensured that, within a given treatment, no one ever met the same person more than once. In every period, the group members knew nothing about the previous cooperation and punishment decisions of the others in the

group, which ensured that subjects could not develop any kind of reputation. At the end of each period, subjects were informed about their own decisions, the decisions of the other group members, and their monetary pay-off in the current period.

At the beginning of the experiment, subjects received written instructions (available from the authors on request) that explained to them the pay-off structure of the game, the random recombination of groups across periods, the anonymous identities of the other session members, the undisclosed history of the previous actions of their current group members, and the fact that they would be privately paid their experimental earnings at the end of the session. After subjects had read the instructions, but before the start of the experiment, we determined whether subjects understood the pay-off structure of the game: subjects had to compute their own pay-off and the pay-offs of their group members in several hypothetical examples. Every subject solved these exercises correctly. All experimental decisions were made on a computer screen using the experimental software z-tree²⁸. Each of the 24 computers was located in a booth such that subjects could not see each other. To relate subjects' decisions to their dispositions, subjects filled out a personality questionnaire before and after the experiment. In addition, subjects filled out an emotions questionnaire after the experiment.

Received 5 October; accepted 5 November 2001.

- Smuts, B. B., Cheney, D. L., Seyfarth, R. M., Wrangham, R. W., & Struhsaker, T. T. (eds) *Primate Societies* (Univ. Chicago Press, Chicago, 1987).
- Richerson, P. & Boyd, R. in *Ideology, Warfare and Indoctrinability* (eds Eibl-Eibesfeldt, I. & Salter, F.) 71–95 (Berghen Books, New York, 1998).
- Sober, E. & Wilson, D. S. *Unto Others: The Evolution and Psychology of Unselfish Behaviour* (Harvard Univ. Press, Cambridge, Massachusetts, 1998).
- Boyd, R. & Richerson, P. in *Evolution and Culture* (ed. Levinson, S.) (MIT Press, Cambridge, Massachusetts, in the press).
- Hamilton, W. D. Genetical evolution of social behavior I and II. *J. Theor. Biol.* **7**, 1–52 (1964).
- Trivers, R. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).
- Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
- Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984).
- Nowak, M. A., May, R. M. & Sigmund, K. The arithmetics of mutual help. *Sci. Am.* **272**, 76–81 (1995).
- Alexander, R. D. *The Biology of Moral Systems* (Aldine de Gruyter, New York, 1987).
- Nowak, M. A. & Sigmund, K. The dynamics of indirect reciprocity. *J. Theor. Biol.* **194**, 561–574 (1998).
- Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577 (1998).
- Lotem, A., Fishman, M. A. & Stone, L. Evolution of cooperation between individuals. *Nature* **400**, 226–227 (1999).
- Wedekind, C. & Milinski, M. Cooperation through image scoring in humans. *Science* **288**, 850–852 (2000).

- Leimar, O. & Hammerstein, P. Evolution of cooperation through indirect reciprocity. *Proc. R. Soc. Lond. B* **268**, 745–753 (2001).
- Zahavi, A. The cost of honesty (further remarks on the handicap principle). *J. Theor. Biol.* **67**, 603–605 (1977).
- Zahavi, A. Altruism as a handicap—the limitations of kin selection and reciprocity. *J. Avian. Biol.* **26**, 1–3 (1995).
- Gintis, H., Smith, E. & Bowles, S. Costly signalling and cooperation. *J. Theor. Biol.* **213**, 103–119 (2001).
- Clutton-Brock, T. H. & Parker, G. A. Punishment in animal societies. *Nature* **373**, 209–216 (1995).
- Axelrod, R. An evolutionary approach to norms. *Am. Pol. Sci. Rev.* **80**, 1095–1111 (1986).
- Heckathorn, D. D. Collective action and the second-order free-rider problem. *Ration. Soc.* **1**, 78–100 (1989).
- Henrich, J. & Boyd, R. Why people punish defectors. *J. Theor. Biol.* **208**, 79–89 (2001).
- Ostrom, E., Walker, J. & Gardner, R. Covenants with and without a sword: self governance is possible. *Am. Pol. Sci. Rev.* **86**, 404–417 (1992).
- Hirschleifer, J. in *The Latest on the Best: Essays on Evolution and Optimality* (ed. Dupré, J.) (MIT Press, Cambridge, Massachusetts, 1987).
- Frank, R. *Passions within Reason: The Strategic Role of the Emotions* (Norton, New York, 1988).
- Gintis, H. Strong reciprocity and human sociality. *J. Theor. Biol.* **206**, 169–179 (2000).
- Sigmund, K., Hauert, C. & Nowak, M. A. Reward and punishment. *Proc. Natl Acad. Sci. USA* **98**, 10757–10761 (2001).
- Fischbacher, U. *Z-tree: Zürich Toolbox for Readymade Economic Experiments* Working Paper No. 21 (Institute for Empirical Research in Economics, Univ. Zürich, 1999).

Acknowledgements

Support by the MacArthur Foundation Network on Economic Environments and the Evolution of Individual Preferences and Social Norms, and the EU-TMR Research Network ENDEAR is gratefully acknowledged. We also thank R. Boyd, A. Falk, U. Fischbacher, H. Gintis and J. Henrich for comments, and M. Näf and D. Reding for research assistance. We are particularly grateful to U. Fischbacher for writing the computer software.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.F. (e-mail: efehr@iew.unizh.ch).

Functional organization of the yeast proteome by systematic analysis of protein complexes

Anne-Claude Gavin*, Markus Bösch*, Roland Krause*, Paola Grandi*, Martina Marzioch*, Andreas Bauer*, Jörg Schultz*, Jens M. Rick*, Anne-Marie Michon*, Cristina-Maria Cruciata*, Marita Remor*, Christian Höfert*, Malgorzata Schelder*, Miro Brajenovic*, Heinz Ruffner*, Alejandro Merino*, Karin Klein*, Manuela Hudak*, David Dickson*, Tatjana Rudi*, Volker Gnau*, Angela Bauch*, Sonja Bastuck*, Bettina Huhse*, Christina Leutwein*, Marie-Anne Heurtier*, Richard R. Copley†, Angela Edelmann*, Erich Querfurth*, Vladimir Rybin*, Gerard Drewes*, Manfred Raida*, Tewis Bouwmeester*, Peer Bork†, Bertrand Seraphin†‡, Bernhard Kuster*, Gitte Neubauer* & Giulio Superti-Furga*†

* Cellzome AG, Meyerhofstrasse 1, 69117 Heidelberg, Germany

† European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

‡ CGM-CNRS, 91198 Gif sur Yvette Cedex, France

Most cellular processes are carried out by multiprotein complexes. The identification and analysis of their components provides insight into how the ensemble of expressed proteins (proteome) is organized into functional units. We used tandem-affinity purification (TAP) and mass spectrometry in a large-scale approach to characterize multiprotein complexes in *Saccharomyces cerevisiae*. We processed 1,739 genes, including 1,143 human orthologues of relevance to human biology, and purified 589 protein assemblies. Bioinformatic analysis of these assemblies defined 232 distinct multiprotein complexes and proposed new cellular roles for 344 proteins, including 231 proteins with no previous functional annotation. Comparison of yeast and human complexes showed that conservation across species extends from single proteins to their molecular environment. Our analysis provides an outline of the eukaryotic proteome as a network of protein complexes at a level of organization beyond binary interactions. This higher-order map contains fundamental biological information and offers the context for a more reasoned and informed approach to drug discovery.

A formidable challenge of postgenomic biology is to understand how genetic information results in the concerted action of gene products in time and space to generate function. In medicine, this is perhaps best reflected in the numerous disorders based on polygenic traits and the notion that the number of human diseases exceeds the number of genes in the genome¹. Moreover, the total number of human genes does not differ substantially from the number of genes of the nematode worm *Caenorhabditis elegans*, suggesting that ‘complexity’ may partly rely on the contextual combination of the gene products². Dissecting the genetic and biochemical circuitry of a cell is a fundamental problem in biology. At the biochemical level, proteins rarely act alone; rather, they interact with other proteins to perform particular cellular tasks³. These assemblies represent more than the sum of their parts by having a new ‘function’³.

Our knowledge regarding the identity of the building elements of specific complexes is limited and is based on selected biochemical approaches and genetic analyses. The only comprehensive protein-interaction studies are based on *ex vivo* and *in vitro* systems, such as two-hybrid systems^{4–6} and protein chips⁷, and need to be integrated with more-physiological approaches. Whenever it has been possible to retrieve and analyse particular cellular protein complexes under physiological conditions, the insight gained from the analysis has been fundamental for the biological understanding of their function, and has often taken the analysis well beyond the limits of genetic analysis^{8,9}. Prominent examples are the spliceosome, the cyclosoome, the proteasome, the nuclear pore complex and the synaptosome^{10–16}. No systematic analysis of protein complexes from the same cell type using the same technique has yet been reported. We have performed a comprehensive analysis of protein complexes of baker’s yeast, *S. cerevisiae*, a model system relevant to human biology^{17,18}.

Large-scale analysis of protein complexes

To systematically purify multiprotein complexes, we developed the strategy depicted in Fig. 1. Gene-specific cassettes containing the TAP tag¹⁹, generated by polymerase chain reaction (PCR), were inserted by homologous recombination at the 3′ end of the genes. We processed 1,739 genes, including 1,143 genes representing eukaryotic orthologues². Orthologues are thought to have evolved by vertical descent from a common ancestor²⁰ and are presumed to carry out the same function. For comparison, we also targeted a nonorthologous set of 596 genes from chromosomes 1, 2 and 4. To test the tagged genes in the absence of the wild-type allele, we used haploid cells. We generated a library of 1,548 yeast strains, of which 1,167 expressed the tagged proteins to detectable levels (Fig. 1). After growing cells to mid-log phase, assemblies were purified from total cellular lysates by TAP¹⁹. This technique combines a first high-affinity purification, mild elution using a site-specific protease, and a second affinity purification to obtain protein complexes with high efficiency and specificity²¹. The purified protein assemblies were separated by denaturing gel electrophoresis, individual bands were digested by trypsin, analysed by matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry (MALDI–TOF MS)⁹ and identified by database search algorithms. In all, 293 proteins were localized at membranes (integral and peripherally associated)²². Because their purification required a separate protocol, only 70 of the membrane-associated proteins were analysed, of which 40 were purified successfully. We analysed the proteins found associated to the 589 (418 orthologues) successfully purified tagged proteins (the ‘raw’ set of purifications; see Supplementary Information Table S1). This generated 20,946 samples for mass spectrometry and subsequently identified 16,830 proteins. Of these, 1,440 were distinct gene products, representing about 25% of the open reading frames (ORFs) in the genome. The

analysis covers proteins of various subcellular compartments, supporting the generality of our approach (Figs 1b, 2a).

Sensitivity, specificity and reliability of the approach

Of the 589 purified tagged proteins, 78% presented associated partners, showing that the method is very efficient for the large-scale retrieval and identification of cellular protein complexes. There are several possible reasons why, in the remaining 22%, we were unable to purify and identify interacting proteins. Particular proteins may not form any or sufficiently stable or soluble

complexes. In other cases, the 20K (relative molecular mass M_r 20,000) TAP tag may interfere with complex assembly or protein localization and function. Because we used haploid cells, we were able to score for viability. In 18% of the cases when essential genes were tagged, we did not obtain viable strains, confirming that carboxy-terminal tagging can impair protein function. Furthermore, the method may fail to detect transient interactions, low stoichiometric complexes, and/or those interactions occurring only in specific physiological states not present or under-represented in exponentially growing cells. In addition, the size distribution of the

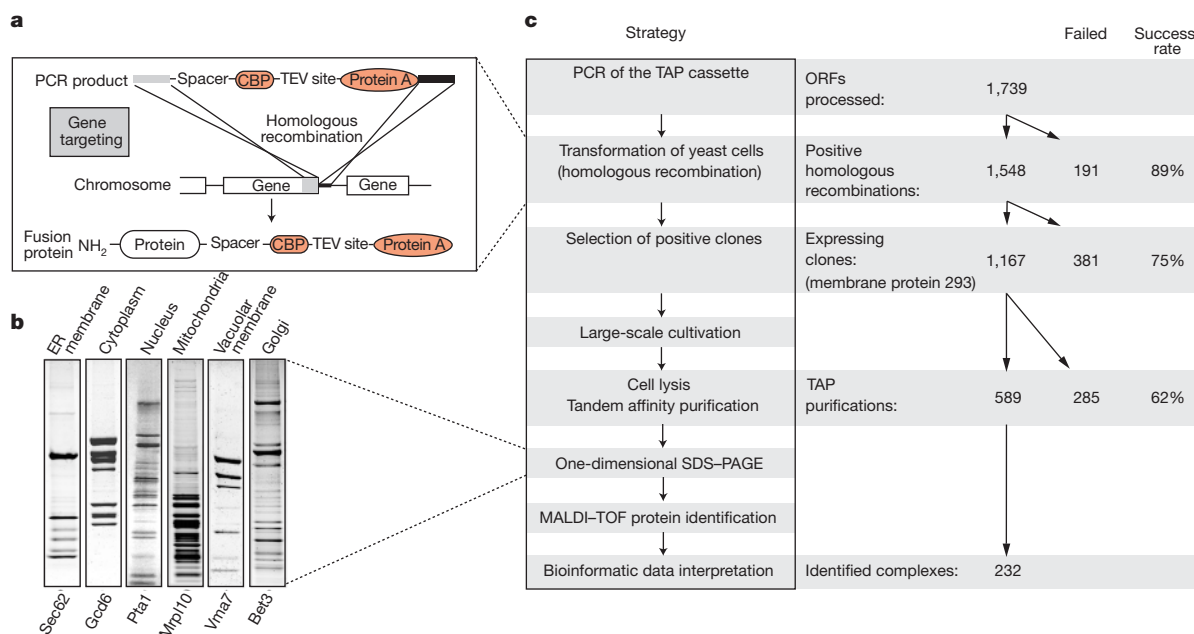


Figure 1 Synopsis of the screen. **a**, Schematic representation of the gene targeting procedure. The TAP cassette is inserted at the C terminus of a given yeast ORF by homologous recombination, generating the TAP-tagged fusion protein. **b**, Examples of TAP complexes purified from different subcellular compartments separated on denaturing

protein gels and stained with Coomassie. Tagged proteins are indicated at the bottom. ER, endoplasmic reticulum. **c**, Schematic representation of the sequential steps used for the purification and identification of TAP complexes (left), and the number of experiments and success rate at each step of the procedure (right).

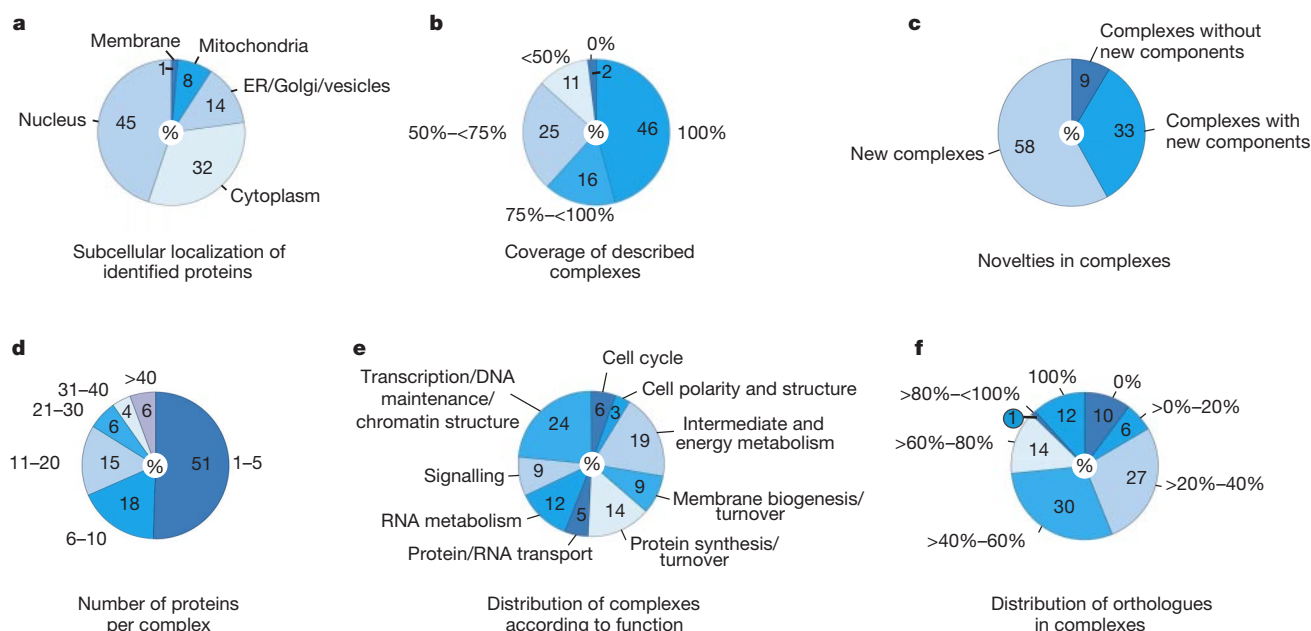


Figure 2 Statistics of proteins and complexes. Numbers inside pie charts represent the percentages of total proteins (**a**) and complexes (**b–f**). Outer labels show partitioning of the data according to the chart function.

identified proteins reveals a clear technical bias against proteins below 15K (see Supplementary Information Fig. S1). However, in about 30% of the cases where we failed to purify complexes around a given protein, the protein was detected when other complex components (entry points) were tagged and purified.

To assess the quality of the results obtained for purifications that do contain associated proteins, we compared our data to the literature (see below). We also established reproducibility of the approach by purifying 13 large complexes at least twice. The probability of detecting the same protein in two different purifications from the same entry point is about 70%. Therefore, on average, 30% of all detected associations presented in this study need to be treated with caution. This is particularly the case if complexes were retrieved from only one entry point. This variability may be inherent to the technique (biological samples, purification, mass spectrometry) but also to the large-scale nature of the approach, tuned to high complex coverage and high sensitivity.

To determine the experimental background, we purified mock-transformed control strains, leading to the identification of 17 contaminant proteins, mainly heat-shock and ribosomal proteins (see Supplementary Information Table S2). These are highly expressed proteins²³. Because these proteins appeared in more than 20 of our purifications (3.5%), we decided to use this cutoff point and pragmatically excluded another 49 proteins present at equal or higher frequencies in our screen from further analysis (see Supplementary Information Table S2). It is prudent to interpret data concerning often-purified proteins just below this cutoff point with caution. Finally, we cannot exclude the occasional artificial interaction generated during cell lysis.

Although stoichiometry was not assessed in this study, we have observed that proteins belonging to the contaminant list generally comprise the weak bands after staining, and typically are identified by fewer tryptic peptides. Moreover, the systematic cutting and analysis of gel slices led to the identification by mass spectrometry even of invisible proteins. The sensitivity of the TAP MS method is high, because we were able to identify proteins present at 15 copies per cell (data not shown). Identified proteins were 6.6K to 559K in size and ranged in pI between 3.9 and 12.4 (see full evaluation in Supplementary Information Fig. S1).

Organization of the purified assemblies into complexes

On the basis of substantial overlaps, we grouped the biochemical purifications obtained with 589 different entry points into a reduced number of biologically meaningful complexes. A total of 245 purifications corresponded to 98 known nonredundant multi-protein complexes present in the yeast protein database (YPD; 60% of the data set)²². A further 242 purifications were assembled into 134 new complexes. The remaining 102 proteins showed no detectable association with other proteins when purified directly, or as part of other complexes. The subsequent statistical analysis is based on a list that includes 232 annotated 'TAP complexes' (see Supplementary Information Table S3).

Among the complexes that were assigned to the known YPD complexes, coverage of components was very high (Fig. 2b). Of all 232 TAP complexes, only 9% had no novel element (Fig. 2c). The size of the TAP complexes varied from 2 to 83 components, with an average of 12 components per complex (Fig. 2d). We assigned cellular roles to complexes by computing functional assignments of the individual components according to YPD²² and by literature mining (Fig. 2e, Supplementary Information Table S3). In general terms, there seemed to be a wide functional distribution of complexes over nine categories (Fig. 2e). Of the 304 proteins with no YPD functional annotation that were identified in our screen, we propose roles for 231 (Supplementary Information Table S3). Moreover, for 113 proteins that had a functional annotation, we discovered a new molecular context (Supplementary Information Tables S3, S4; see also examples below).

Cohesive and dynamic complexes

A particular complex is not necessarily of invariable composition nor are all its building blocks uniquely associated with that specific complex. With several distinct tagged proteins as entry points to purify a complex, core components can be identified and validated, whereas more dynamic, perhaps regulatory components may be present differentially. The dynamics of complex composition are well illustrated by the cellular signalling complexes formed around the protein phosphatase 2A (PP2A; yeast TAP-C151; see Supplementary Information Table S3). Tagging different known PP2A components resulted in the purification of the known trimeric complexes containing Tpd3 (the regulatory A subunit), either of the two catalytic subunits, Pph21 and Pph22, and either of the two regulatory B subunits, Cdc55 and Rts1. The Cdc55-containing complexes were found to additionally contain Zds1 or Zds2, known cell-cycle regulators, revealing preferences among the different complexes and a link to cell-cycle checkpoints. Additional plasticity of the PP2A complexes is apparent by the interaction with three proteins implicated in bud shape and morphogenesis (Lte1, Kel1 and YBL104C). This analysis also shows that the interactions of a signalling enzyme may be sufficiently strong to allow the detection of distinct cellular complexes and thus be diagnostic for a role of this enzyme in different cellular activities.

An example of a large, cohesive complex is given by the polyadenylation machinery, which is responsible for the sequential steps necessary for eukaryotic messenger RNA cleavage and polyadenylation²⁴ (yeast TAP-C162; see Supplementary Information Table S3). Using Pta1 as the entry point, we identified 12 of the 13 known interactors and 7 new components (Fig. 3a). The

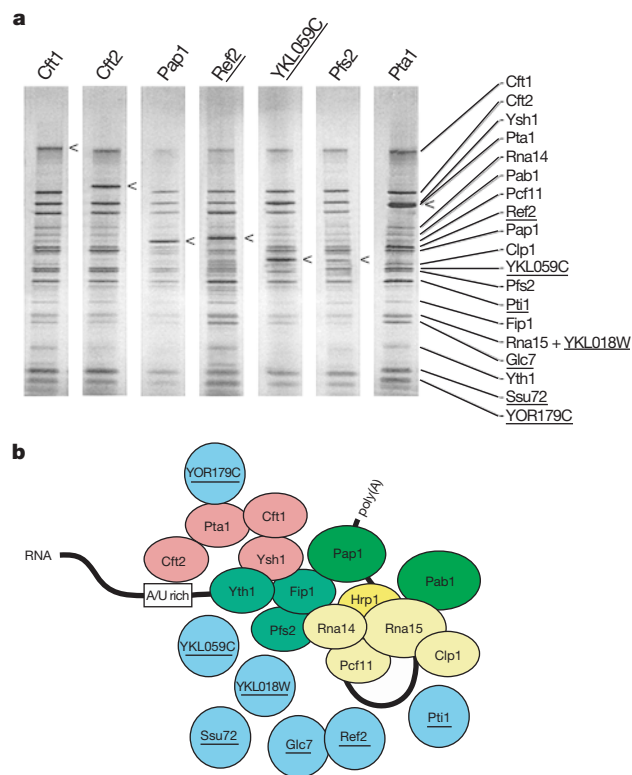


Figure 3 Primary validation of complex composition by 'reverse' purification: the polyadenylation machinery. **a**, A similar band pattern is observed when different components of the polyadenylation machinery complex are used as entry points for affinity purification. Underlined are new components of the polyadenylation machinery complex for which a physical association has not yet been described. The bands of the tagged proteins are indicated by arrowheads. **b**, Proposed model of the polyadenylation machinery.

composition of the complex was validated by extensive 'reverse' analysis, including purifications obtained with YKL059C and Ref2, two of the new interactors (Fig. 3a). Four new components—Pti1, YKL018W, YKL059C and YOR179C—had no previous functional annotation, but could be included into a hypothetical model by bioinformatic analysis (Fig. 3b). Ssu72, another of the seven new interactors, had previously been reported to interact with TFIIB, strongly supporting a suspected link between the polyadenylation machinery and transcription. Thus, complexes are often sufficiently strong to show high composition integrity even when purified with different entry points.

A higher-order organization map of the proteome

After assigning individual proteins to protein complexes, we investigated relationships between complexes to understand the integration and coordination of cellular functions. We represented relationship by linking complexes that share components (Fig. 4). By plotting all the relationships, we obtained a network of complexes. Connections in this network not only reflect physical interaction of complexes, but may also represent common regulation, localization, turnover or architecture. Most complexes are linked. The more connected a complex, the more central its position

in the network. Complexes composed of at least 50% orthologues are shown as double sized, and complexes are colour-coded according to cellular roles. Several complexes belonging to the same class appear to group, suggesting that sharing of components reflects functional relationships. These relationships are best observed with complexes involved in mRNA metabolism (orange), cell cycle (red), protein synthesis and turnover (light green), and intermediate and energy metabolism (violet). Complexes involved in transport of protein or RNA (pink), in contrast, appear more dispersed and have connections to complexes of all other cellular roles. There are several 'satellite' complexes that do not seem to share components. Because this analysis is not exhaustive, we expect more connections for some of these complexes as more are purified and analysed. A software package (available at <http://yeast.cellzome.com>) allows the navigation of this proteome map at both the protein and complex level. Such a tool is essential to allow for proper data interpretation and for the generation of hypotheses leading to further experimental investigations.

Parallel analysis of human and yeast complexes

Orthologous gene products are thought to be responsible for essential cellular activities. We found that orthologous proteins

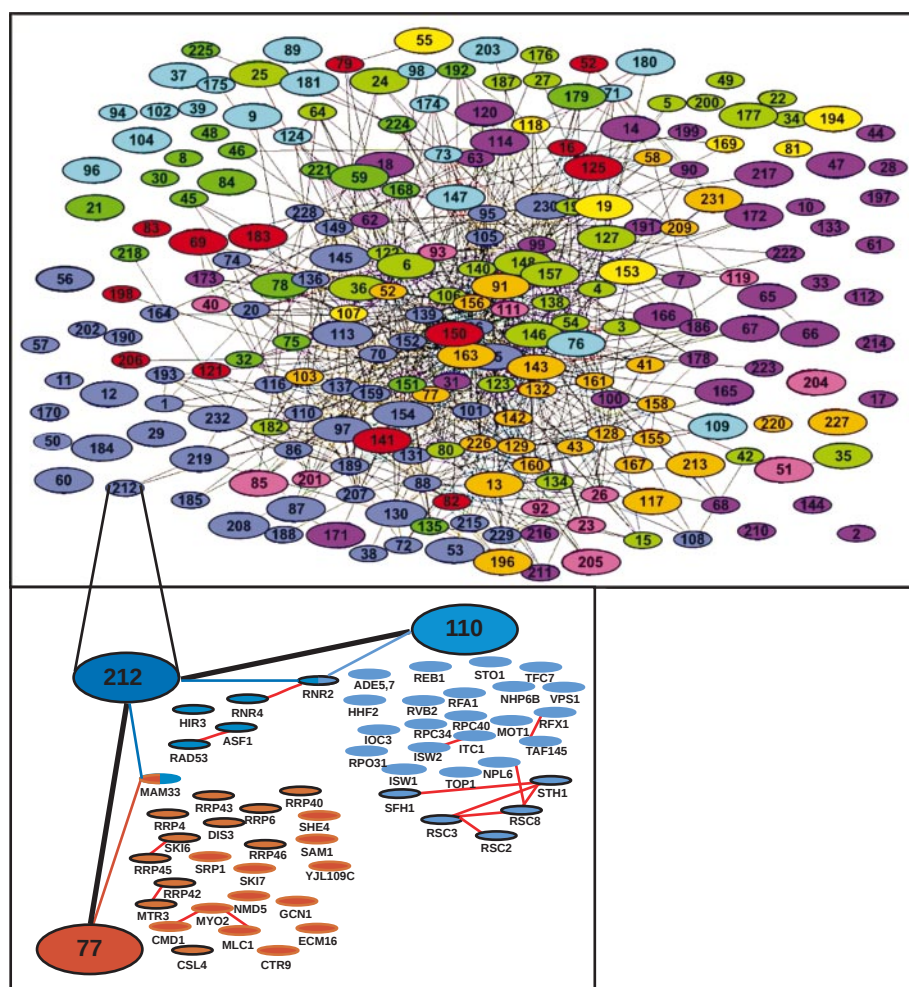


Figure 4 The protein complex network, and grouping of connected complexes. Links were established between complexes sharing at least one protein. For clarity, proteins found in more than nine complexes were omitted. The graphs were generated automatically by a relaxation algorithm that finds a local minimum in the distribution of nodes by minimizing the distance of connected nodes and maximizing distance of unconnected nodes. In the upper panel, cellular roles of the individual complexes (ascribed in Supplementary Information Table S3) are colour coded: red, cell cycle; dark green, signalling; dark blue,

transcription, DNA maintenance, chromatin structure; pink, protein and RNA transport; orange, RNA metabolism; light green, protein synthesis and turnover; brown, cell polarity and structure; violet, intermediate and energy metabolism; light blue, membrane biogenesis and traffic. The lower panel is an example of a complex (yeast TAP-C212) linked to two other complexes (yeast TAP-C77 and TAP-C110) by shared components. It illustrates the connection between the protein and complex levels of organization. Red lines indicate physical interactions as listed in YPD²².

preferentially interact with complexes enriched with other orthologues (mean 53%; Fig. 2f). In comparison, nonorthologous proteins have a lower propensity for such interaction (mean 31%). Similarly, the likelihood to interact with essential gene products is higher for essential (44%) than for nonessential (17%) proteins. This supports the existence of an 'orthologous proteome' that may represent core functions for the eukaryotic cell¹⁸. To determine whether the TAP strategy can be applied to retrieve equivalent multiprotein complexes from yeast and human cells, we compared different complexes from distinct subcellular compartments: Arp2/3, a cytoskeleton-associated complex, Ccr4–Not, a nuclear assembly, and Trapp, a Golgi-associated complex. The Arp2/3 complex is a stable multiprotein assembly required for the nucleation of actin filaments in all eukaryotic cells and consists of seven proteins in human and yeast²⁵. TAP of Arp2 in yeast (TAP-C153) and ARPC2 in human²⁵ resulted in the isolation and identification of all known components. This indicates that the TAP approach combined with liquid chromatography coupled to tandem mass spectrometry (LC/MS/MS) is an efficient and sensitive method for the retrieval and characterization of human multiprotein complexes (Fig. 5a).

The yeast Ccr4–Not complex (TAP-C149) is involved in the control of gene expression and consists of eight components²⁶. Potential human orthologues have been identified and characterized for yeast Not2, Not3, Not4 and Caf1, but not for yeast Not1 and Ccr4. Moreover, no respective multiprotein complex has yet been described in mammalian cells²⁷. TAP of tagged human NOT2 resulted in the identification of a multiprotein complex consisting of human NOT2, CAF1 and CALIF, and two functionally non-annotated gene products, encoded by KIAA1007 and KIAA1194, which we could assign as the orthologues of yeast Not1 and Ccr4, respectively (Fig. 5b). Purification of tagged yeast Ccr4 resulted in the identification of a complex component, Caf40, that has an orthologous counterpart, Rqcd1, also identified in the human complex. These data strongly suggest that the human and yeast Ccr4–Not complexes are comparable in subunit composition.

As a third example we purified and characterized an orthologous human TRAPP (transport protein particle) complex. The yeast complex contains ten subunits that are required for docking of transport vesicles derived from the endoplasmic reticulum to the *cis*-Golgi (yeast TAP-C102)²⁸. The human complex had been purified previously as an assembly of about 670K; however, apart from human BET3 and TRS20, none of the other complex subunits had been identified²⁸. TAP purification of tagged human BET3 resulted in the identification of a complex consisting of human BET3, MUM2, R32611_2, Sedlin, EHOC-1, PTD009 and KIAA1012, which we assigned as the orthologues of yeast Bet3, Bet5, Trs33, Trs20, Trs130, Trs23 and Trs85, respectively (Fig. 5c). Taken together, these examples show that the analysis of yeast complexes can often predict the composition of the human counterparts. This large-scale yeast proteome analysis could have immediate functional implications for human biology.

Discussion

To assign cellular functions to new, nonannotated gene products, and to understand the context in which proteins operate, several large-scale approaches have been undertaken. These include monitoring of mRNA expression (chips and serial analysis of gene expression (SAGE))²⁹, loss-of-function approaches combined with subcellular localization screens (in yeast^{29,30}, RNA-mediated interference in *C. elegans*^{31,32}, gene trap in mice^{33,34}), computational *in silico* methods (protein fusions, gene neighbouring, structural predictions)^{35,36}, and extensive two-hybrid screens^{4–6} and protein chip analysis^{7,37}. The TAP/MS-based functional proteomics approach presented here may well constitute the largest analysis of protein complexes to date. We confirmed the expression of 1,440 ORFs as annotated in the genome, of which 59 had been assigned

only as hypothetical. A large-scale analysis of yeast proteins performed previously on a crude cell extract identified 1,484 different proteins from exponentially growing cells³⁸. Another 714 proteins were detected in our study. This raises the total number of ascertained proteome components to 2,210.

TAP¹⁹ proved invaluable for the purification of complexes from different cellular compartments, including complexes associated with cellular membranes. The approach also allows for the efficient identification of low-abundance proteins that would not be detectable by approaches involving expression proteomics^{9,21}. Further,

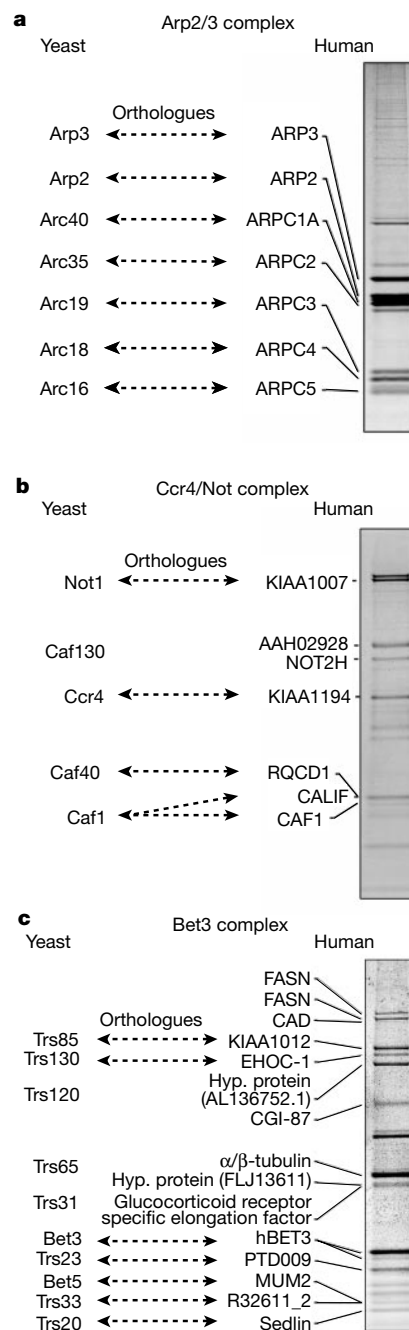


Figure 5 Protein complexes have a similar composition in yeast and human. Comparison of three TAP protein complexes isolated from human and yeast cells. All orthologous pairs are indicated by arrows, demonstrating that the complex composition between yeast and human is largely conserved. Coomassie-stained gels are shown only for the human purifications. **a**, Arp2/3 complex; **b**, Ccr4–Not2 complex; **c**, Trapp complex. Hyp. protein, hypothetical protein.

TAP allows the purification of very large complexes. For example, we were able to purify yeast TAP-C116 (Ino80), a complex reported to have an M_r about 1M–1.5M (ref. 39), and we identified all known and several new interactors.

We sought to gauge the reliability of our data by comparing the experimental results to the literature. Although comparison of our data to the YPD complex database is straightforward (Fig. 2b), it is very difficult to derive meaningful comparative information from the yeast two-hybrid data because the latter produces binary interactions, whereas the TAP/MS method yields complex composition data. When considering all possible interactions between proteins within a complex or a purification, normalized for the proteins that were identified in this study, we find that our data overlap with only 7% of the interactions seen by yeast two-hybrid assays^{4–6}. When compared with the YPD protein complexes, this study covers 56%, whereas large-scale yeast two-hybrid approaches cover 10%. There are two reasons for this difference. On one hand, the yeast two-hybrid approaches touch only 35% of the described complexes, compared with 60% in this study. On the other hand, there is a generally lower coverage of individual components within complexes. The figure for the yeast two-hybrid data is 39% compared with 56% in this study. To achieve the respective coverage, the yeast two-hybrid approaches required processing of 95% of all yeast ORFs compared with 25% in our study. Altogether, this illustrates that the two methodologies address different aspects of protein interaction. Comprehensive two-hybrid approaches do not seem to be particularly suited for characterization of protein complexes. This supports the view that complex formation is more than the sum of binary interactions. However, two-hybrid analysis is of exceptional value for the detection of pairwise and transient associations. The success of the TAP/MS approach for the characterization of protein complexes relies on the conditions used for the assembly and retrieval of the complexes. These include maintaining protein concentration, localization and post-translational modifications in a manner that closely approximates normal physiology. Because the TAP/MS method does not provide information on the orientation of complex components, complex characterization and yeast two-hybrid analysis are ideally complementary.

Another outcome of our study is the ease and frequency by which protein complexes can be retrieved from cells. These biophysical properties of protein complexes may suggest cooperative binding. Bridging factors, post-translational modifications, allosteric structural changes, and binding of ions and metabolites can all cooperate to increase the number of short-range interactions between individual proteins in an assembly. Moreover, several proteins with critical regulatory functions are non-globular or intrinsically unstructured⁴⁰. Folding into ordered structures occurs only on binding to other proteins, offering the opportunity of control over the thermodynamics of the binding process. We anticipate that some of the complexes identified in this study will be useful for structural studies.

Although we used only one set of experimental parameters here to grow and maintain cells for the evaluation of complex composition, we will, in the future, systematically modify experimental parameters to evaluate the impact of a changing environment on complex integrity⁴¹. These studies should help to elucidate the dynamics of complex assembly and disassembly. Moreover, the strains generated can be used for parallel assessment of protein expression levels. Finally, tandem-affinity-purified complexes from this collection may be a starting point to develop protein chips containing physiological protein complexes⁷ and for assessment of biochemical activity of proteins within their molecular environment⁴².

The statistical analysis of the large-scale yeast approach shows a clear tendency of proteins that are part of the set of metazoan orthologues to bind to other proteins of the same set. Moreover, we

also observed a propensity to associate among the products of essential genes. Complexes containing orthologues and essential proteins overlap significantly. This recalls the proposition that the products of essential genes are also more likely to represent central components in a protein network⁴³. Together, this raises the possibility that orthologue complexes represent the building blocks of a eukaryotic 'core proteome' covering basic cellular functions^{18,43}. We believe that a significant number of the yeast complexes described here will have human equivalents and expect that these may form the basis for understanding multifactorial diseases. Through the 'guilt by association' concept, we are able to propose cellular roles for proteins that had no previous functional annotation and new roles for known proteins. Assessment of the physiological molecular context of proteins, as described here, may be one of the most efficient and unambiguous routes towards the assignment of gene identity and function.

Our analysis allowed us to group cellular proteins into about 200 complexes. These complexes are connected to each other by shared components. The network that resulted is a functional description of the eukaryotic proteome at a higher level of organization. Such higher-order maps will bring an increasing quality to our appreciation of biological systems. It is expected that this may provide drug discovery programmes with a molecular context for the choice and evaluation of drug targets. □

Methods

Yeast strain construction and TAP

Yeast strains expressing TAP-tagged ORFs were constructed in a semi-automated way essentially as done previously^{19,21}. Cells were cultured at 30 °C in YPD medium, collected during exponential growth, and lysed mechanically with glass beads. Purifications were done as described^{19,21}. For the purification of membrane proteins, the detergent concentration was adjusted to 1.5% after lysis.

TAP from human cells

Retroviral transduction vectors were generated by directional cloning of PCR-amplified ORFs into a modified version of a MoMLV-based vector via the Gateway site-specific recombination system (Life Technologies). For NOT2 and ARPC2, the TAP cassette was fused to the amino terminus and for BET3 to the C terminus. In all 3 cases, cell cultures were generated by retrovirus-mediated gene transfer and complexes were purified after cell expansion and cultivation for at least 5 days by a modified TAP protocol¹⁹.

High-throughput protein identification

Purified proteins were concentrated, separated on 4–12% NuPAGE gels (Novex) and stained with colloidal Coomassie blue. Gels were sliced into 1.25-mm bands across the entire separation range of each lane to sample all potential interacting proteins without bias with respect to size and relative abundance. Cut bands were digested with trypsin essentially as described⁴⁴. The resulting tryptic peptide mixtures were analysed by automated MALDI-TOF MS (Voyager DE-STR, Applied Biosystems). Proteins were identified by automated peptide mass fingerprinting using the software tool Knexus (Proteometrics) and an in-house built sequence database of *S. cerevisiae* proteins. Experiments with human orthologues of yeast proteins were subjected to the same cutting and digestion procedure but protein identification was accomplished by automated LC/MS/MS analysis (Ultimate, LC Packings, QTOF2, Micromass) in conjunction with searches of the GenPept database (<ftp://ftp.ncbi.nlm.nih.gov/genbank/genpept.fsa.gz>) using the software tool Mascot (Matrix Science).

Bioinformatics

Functional and localization information about yeast proteins was retrieved from the YPD released in August 2001. To get a more concise classification for localization and function, YPD classes were merged. For the analysis of assembled complexes from our purifications to the published complexes, we removed redundancy from the YPD data set manually. Protein domain analysis was performed with SMART⁴⁵. PsiBlast⁴⁶ was used for homology analysis. All additional analysis software was developed by us, using Perl and Python. For the comparison of purifications to the YPD complexes, each member of a complex and, independently, of the purifications, was considered to be connected to each other. Redundant interactions were merged. For the comparison with yeast two-hybrid assays we counted every described binary interaction included in our purifications and included in the YPD complexes, respectively. For calculation of coverage of complexes, we considered only the YPD complexes that included at least two identified components (from yeast two-hybrid interactions or from complex purification). For the calculation of coverage of complex components, all purifications or yeast two-hybrid pairs with an entry point or bait falling into a described complex were considered.

Received 15 August; accepted 25 October 2001.

1. Roses, A. D. Pharmacogenetics and the practice of medicine. *Nature* **405**, 857–865 (2000).
2. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
3. Alberts, B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291–294 (1998).
4. Fromont-Racine, M., Rain, J. C. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* **16**, 277–282 (1997).
5. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
6. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
7. Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105 (2001).
8. Blackstock, W. P. & Weir, M. P. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* **17**, 121–127 (1999).
9. Pandey, A. & Mann, M. Proteomics to study genes and genomes. *Nature* **405**, 837–846 (2000).
10. Neubauer, G. *et al.* Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nature Genet.* **20**, 46–50 (1998).
11. Zachariae, W., Shin, T. H., Galova, M., Obermaier, B. & Nasmyth, K. Identification of subunits of the anaphase-promoting complex of *Saccharomyces cerevisiae*. *Science* **274**, 1201–1204 (1996).
12. Varga-Weisz, P. D. *et al.* Chromatin-remodelling factor CHRAC contains the ATPases ISW1 and topoisomerase II. *Nature* **388**, 598–602 (1997).
13. Verma, R. *et al.* Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol. Biol. Cell* **11**, 3425–3439 (2000).
14. Neubauer, G. *et al.* Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl Acad. Sci. USA* **94**, 385–390 (1997).
15. Rout, M. P. *et al.* The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635–651 (2000).
16. Husi, H., Ward, M. A., Choudhary, J. S., Blackstock, W. P. & Grant, S. G. Proteomic analysis of NMDA receptor-adhesion protein signaling complexes. *Nature Neurosci.* **3**, 661–669 (2000).
17. Bassett, D. E. Jr, Boguski, M. S. & Hieter, P. Yeast genes and human disease. *Nature* **379**, 589–590 (1996).
18. Rubin, G. M. *et al.* Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215 (2000).
19. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
20. Fitch, W. M. Distinguishing homologous from analogous proteins. *Syst. Zool.* **19**, 99–113 (1970).
21. Puig, O. *et al.* The tandem affinity purification (tap) method: a general procedure of protein complex purification. *Methods* **24**, 218–229 (2001).
22. Costanzo, M. C. *et al.* YPD, PombePD and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information. *Nucleic Acids Res.* **29**, 75–79 (2001).
23. Garrels, J. I. *et al.* Proteome studies of *Saccharomyces cerevisiae*: identification and characterization of abundant proteins. *Electrophoresis* **18**, 1347–1360 (1997).
24. Barabino, S. M. & Keller, W. Last but not least: regulated poly(A) tail formation. *Cell* **99**, 9–11 (1999).
25. Higgs, H. N. & Pollard, T. D. Regulation of actin filament network formation through Arp2/3 complex: activation by a diverse array of proteins. *Annu. Rev. Biochem.* **70**, 649–676 (2001).
26. Liu, H. Y. *et al.* The NOT proteins are part of the CCR4 transcriptional complex and affect gene expression both positively and negatively. *EMBO J.* **17**, 1096–1106 (1998).
27. Albert, T. K. *et al.* Isolation and characterization of human orthologs of yeast CCR4–NOT complex subunits. *Nucleic Acids Res.* **28**, 809–817 (2000).
28. Sacher, M., Barrowman, J., Schieltz, D., Yates, J. R. III & Ferro-Novick, S. Identification and characterization of five new subunits of TRAPP. *Eur. J. Cell Biol.* **79**, 71–80 (2000).
29. Lockhart, D. J. & Winzler, E. A. Genomics, gene expression and DNA arrays. *Nature* **405**, 827–836 (2000).
30. Ross-Macdonald, P. *et al.* Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**, 413–418 (1999).
31. Fraser, A. G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).
32. Gonczy, P. *et al.* Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331–336 (2000).
33. Friedrich, G. & Soriano, P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* **5**, 1513–1523 (1991).
34. Leighton, P. A. *et al.* Defining brain wiring patterns and mechanisms through gene trapping in mice. *Nature* **410**, 174–179 (2001).
35. Eisenberg, D., Marcotte, E. M., Xenarios, I. & Yeates, T. O. Protein function in the post-genomic era. *Nature* **405**, 823–826 (2000).
36. Huynen, M., Snel, B., Lathe, W. III & Bork, P. Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res.* **10**, 1204–1210 (2000).
37. Zhu, H. *et al.* Analysis of yeast protein kinases using protein chips. *Nature Genet.* **26**, 283–289 (2000).
38. Washburn, M. P., Wolters, D. & Yates, J. R. III Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* **19**, 242–247 (2001).
39. Shen, X., Mizuguchi, G., Hamiche, A. & Wu, C. A chromatin remodelling complex involved in transcription and DNA processing. *Nature* **406**, 541–544 (2000).
40. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. *J. Mol. Biol.* **293**, 321–331 (1999).
41. Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934 (2001).
42. Martzen, M. R. *et al.* A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**, 1153–1155 (1999).
43. Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
44. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
45. Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. & Bork, P. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28**, 231–234 (2000).
46. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank C. Cohen, S. Artavanis-Tsakonas and F. Kafatos for suggestions and support throughout the work; C. Gässler, C. Kalla, N. Umhey and P. Völkel for help with experiments; K. Herzog and her team for media supply; E. Hurt and H. T. Timmers for sharing unpublished results and for suggestions; and numerous people at the European Molecular Biology Laboratory for help, especially G. Bricard, F. Caspary, O. Puig, G. Stier and members of the Nebreda, Superti-Furga, Klein, Wilm and Bork laboratories.

Correspondence and requests for materials should be addressed to A.C.G. (e-mail: anne-claude.gavin@cellzome.com) or G.S.F. (giulio.superti-furga@cellzome.com). The data are accessible at <http://yeast.cellzome.com>.

A faint discrete source origin for the highly ionized iron emission from the Galactic Centre region

Q. D. Wang*, E. V. Gotthelf† & C. C. Lang*

* Department of Astronomy, University of Massachusetts, Amherst, Massachusetts 01003, USA

† Columbia Astrophysics Laboratory, Columbia University, New York, New York 10027, USA

The origin of the X-ray emission for the central region of our Galaxy has remained a mystery^{1–4}. In particular, the relative spectral contributions of the diffuse emission and discrete sources, which are critical to understanding the high-energy phenomena in this environment, have been unclear because of the lack of sufficient spatial resolution. Here we report the results of a large-scale imaging survey of the Galactic Centre that resolves these components. We find that the K α emission from iron that has been highly ionized (so that it has only two electrons left), which has previously been attributed to the diffuse component¹, actually arises mainly from discrete sources. This suggests that the presence of a large amount of hot gas ($T \approx 10^8$ K) is no longer required to explain the iron line emission. The spectra of the discrete sources indicate the presence of numerous accreting white dwarfs, neutron stars, and/or black holes in the region. The diffuse emission dominates over the contribution from the faint point sources, and is shown to be associated globally with interstellar features that have been observed at radio and mid-infrared wavelengths, suggesting that it is the product of recent massive star formation.

We have carried out a systematic X-ray survey of the Galactic Centre region at arcsecond scales with the Chandra X-ray observatory⁵. Figure 1 shows the image of the X-ray emission arising from the Galactic Centre region. The figure, though optimized to reveal low-surface-brightness features, shows many discrete sources. Our preliminary source detection analysis results in $\sim 1,000$ discrete

sources in the field with a signal-to-noise ratio of $\geq 3\sigma$. Less than 20 of these sources are previously known objects, most of which are bright X-ray binaries^{6,7}; examples include the two bright accreting X-ray binaries, 1E1743.1–2843 and 1E1740.7–2942. On the basis of a comparison with the source density in a relatively blank region of the Galactic plane⁸, we estimate that up to half of our newly-detected sources could be luminous background active galactic nuclei, the X-ray radiation from which is highly attenuated by the Galactic interstellar absorption. The X-ray absorption also varies across the field, affecting the surface brightness distribution, particularly in the energy band of 1–3 keV. Most of the detected sources are in the energy range of 2–10 keV, with a luminosity range of 10^{32} – 10^{35} erg s^{–1} at the distance of the Galactic Centre.

We consider the nature of the large number of newly detected X-ray sources by examining their composite spectral signature, and comparing this to the diffuse emission (Fig. 2). The accumulated source spectrum (excluding the four brightest ones) within the central region shows a distinct emission feature centred at ~ 6.7 keV with a gaussian width of ~ 0.09 keV, which agrees with the previous measurements⁴. This spectral feature represents the He-like Fe K α -line triple at 6.64, 6.67 and 6.70 keV, with a possible modest contribution from the 6.4-keV fluorescent line of neutral or mildly ionized Fe atoms. These emission lines are characteristic of X-ray binaries containing white dwarfs, neutron stars or black holes, especially at relatively quiescent states^{9,10}. This X-ray binary explanation is also consistent with the otherwise flat and relatively featureless shape of the spectrum. The weak, narrow emission lines (at ~ 2.4 keV and possibly 3 keV) are possibly due to a relatively small contribution from massive stars. The X-ray spectrum of the luminous Arches star cluster¹¹, for example, shows distinct narrow lines, and is considerably softer than the mean source spectrum shown in Fig. 2. The upturn of the source spectrum below ~ 2 keV is a result of the contribution from relatively nearby sources, which suffers less soft X-ray absorption by the interstellar medium than does the X-ray radiation from the Galactic Centre.

Our survey also shows that large amounts of diffuse X-ray emitting material are distributed asymmetrically around the Galactic Centre, and are particularly concentrated in the region about $10'$ offset from the Sgr A radio complex (ref. 12; to the lower left of Sgr A in Fig. 1). Several additional large-scale diffuse X-ray features are

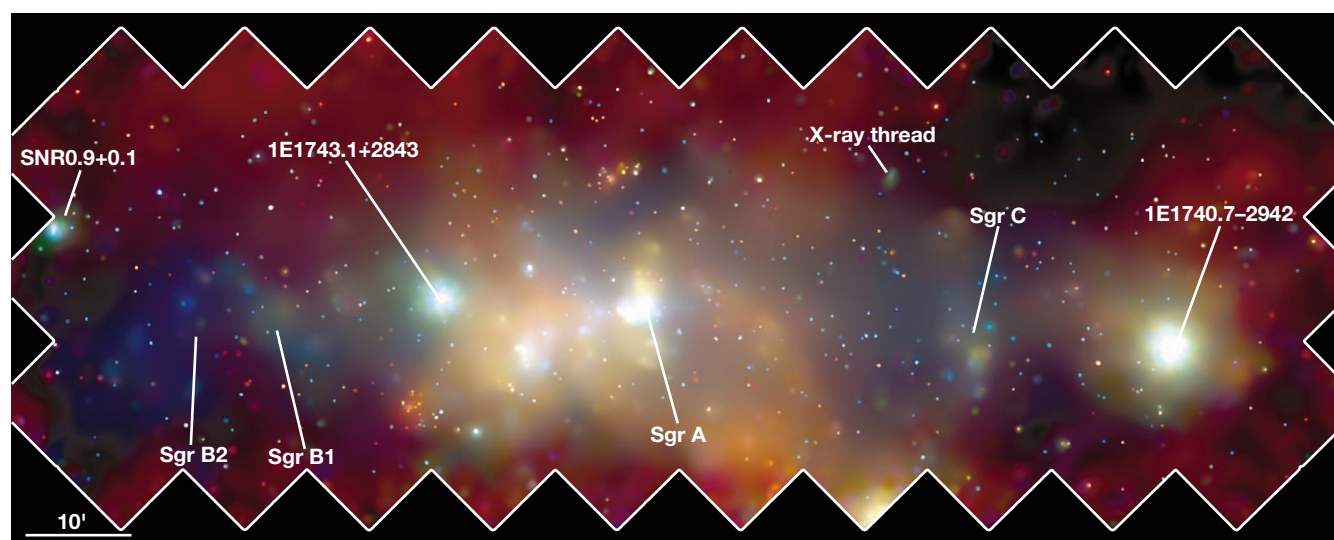


Figure 1 Mosaic image of the Galactic Centre region. This image covers a band of about $2^\circ \times 0.8^\circ$ in Galactic coordinates, and is centred at longitude $l = -0.1^\circ$, latitude $b = 0^\circ$, roughly the location of the Sgr A complex. The three energy bands are 1–3 keV (shown in red), 3–5 keV (green), and 5–8 keV (blue). The data consist of 30 separate pointings, and were acquired during July 2001 with the front-illuminated Advanced CCD Imaging Spectrometer (ACIS-I). The spatial resolution ranges from $\sim 0.5''$ on-axis, to $\sim 5''$

at the near edge of the ACIS-I, and to $\sim 10''$ at the diagonal edge for each pointing. This image is adaptively smoothed with a signal-to-noise ratio of 3. The intensity is plotted logarithmically to emphasize low-surface-brightness features. Standard imaging calibration and exposure correction have been applied, as well as corrections for charge transfer inefficiency effects^{5,25}.

present in the soft band (1–3 keV) around Sgr A*, some of which apparently extend into regions of high Galactic latitude outside the field of view of the X-ray image. There are also prominent features associated with the Sgr C complex and near the Galactic microquasar 1E1740.7–2942. The count rate ratio of the sources to the enhanced diffuse component in the two spectra (Fig. 2) is 0.1 in the 2–10 keV band. Even in the regions of the deepest exposure (~ 120 ks in the Sgr B2 region, including an archival observation¹³), where the source detection limit reaches $\sim 8 \times 10^{31}$ erg s⁻¹, the diffuse emission dominates. There is no known X-ray source population that might contribute substantially below this source detection limit.

The overall shape and line features of the two spectra in Fig. 2 are distinctly different, indicating that the diffuse and point-source components are intrinsically different. In contrast to the spectrum of discrete sources, the diffuse emission spectrum shows distinct, broad spectral features. The He-like Fe K α line (at ~ 6.7 keV) is not as prominent in the diffuse emission spectrum as in the discrete source spectrum. Therefore, the ubiquitous and strong presence of this line in previous X-ray spectra of the Galactic Centre region¹ is primarily due to discrete sources. The reduced strength of the He-like Fe line no longer requires the presence of large amounts of gas at $\sim 10^8$ K, which are very difficult to explain physically⁴. The weaker He-like Fe line, as well as the prominent ion lines at lower energies (for example, Si XIII K α and S XV K α ; Fig. 2) are consistent with an optically thin thermal plasma with a characteristic temperature of $\sim 10^7$ K, comparable to that typically found in young supernova remnants. Such plasma still contributes significantly to the 6.7-keV line. One example is Sgr A East, probably a young supernova remnant, which has an X-ray spectrum characterized by a thermal plasma with a temperature of ~ 2 keV and shows a strong 6.7-keV emission line¹⁴.

Our spectral results also suggest that non-thermal processes contribute significantly to the diffuse X-ray emission, especially at the higher energies. A recent study³ has demonstrated that the inclusion of a non-thermal component in modelling the X-ray background spectrum from the Galactic plane and toward the Galactic Centre significantly reduces the required characteristic

temperature of the plasma component. But the nature and origin of this contribution remain unclear. Nearly half of our detected diffuse emission in the 5–8 keV band is due to the Fe 6.4-keV line (Fig. 2), which results from the filling of a neutral or weakly ionized Fe K-shell vacancy due either to ionization by hard X-ray radiation (>7.1 keV) or to collision with non-relativistic cosmic rays³. We find that the distribution of the line emission is indeed generally correlated with lumpy dense molecular material^{15,16}. However, currently there is not a sufficient population of bright X-ray sources in the Galactic Centre region to produce the distribution of 6.4-keV line fluorescence^{13,17}. One possibility might be that the luminosity of X-ray sources (for example, the Galactic Centre supermassive black hole) varies greatly, and that the currently low luminosity is abnormal. If the luminosity averaged over the light crossing time of the region (a few hundred years) is several orders of magnitude higher, much of the 5–8 keV band diffuse emission may then be explained as the past point-like emission scattered by (or fluoresced from) molecular clouds found in the region^{15,18}.

We also find that the continuum emission in the 4–6 keV range is substantially more uniformly distributed along the central Galactic plane than both the 6.4-keV line and the emission in the lower-energy bands. This continuum emission may be induced partly by non-thermal electrons. However, we find little detailed correlation between diffuse X-ray and radio features in the Galactic Centre region. Of the eight prominent non-thermal filaments known, only one has a direct X-ray counterpart: G359.54+0.18 (ref. 19; marked as ‘X-ray thread’ in Fig. 1). In particular, the X-ray emission is not correlated with the most prominent non-thermal filaments in the Radio arc or with the mid-infrared shells/arcs (see, for example, Fig. 3). This suggests that the bulk of the emission cannot be due to the inverse Compton scattering of the cosmic microwave background or the interstellar infrared radiation off relativistic electrons²⁰. But bremsstrahlung produced by non-relativistic cosmic-ray electrons ($\lesssim 1$ MeV) remains a possibility³.

Such multiwavelength comparisons have provided us with a new perspective of the interplay between various stellar and interstellar components in the Galactic Centre region. The enhanced diffuse

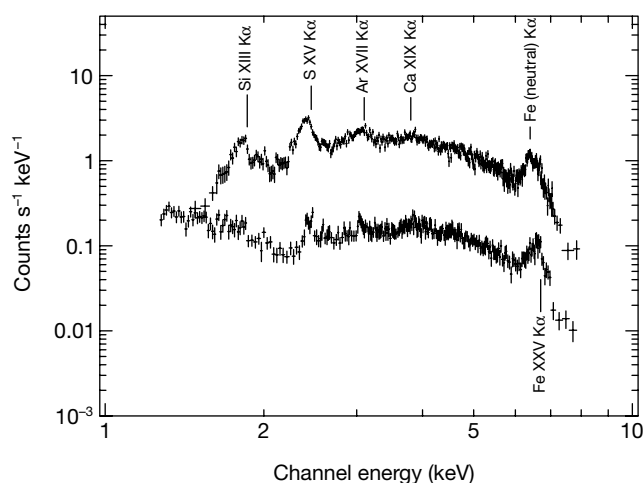


Figure 2 Comparison of accumulated point-source (lower) and diffuse emission (upper) spectra. These two spectra are extracted from an ellipse (with the major and minor axis equal to 50' and 12'), centred at Sgr A* and oriented along the Galactic plane. Regions around the four brightest sources (1E1740.7–2942, 1E1743–2843, 1A1743–288, and the Arches cluster) with count rates greater than 0.1 counts s⁻¹ are excluded to minimize the spectral pile-up problem. The total count rate from these sources is comparable to that of the diffuse emission. For each spectrum, we derive an average auxiliary response, weighted by an image of selected events in the detector coordinates. The response is used for the channel energy scaling and for characterizing the spectral lines. A diffuse background, obtained in the outer field and normalized in both exposure and area, is subtracted from the two spectra.

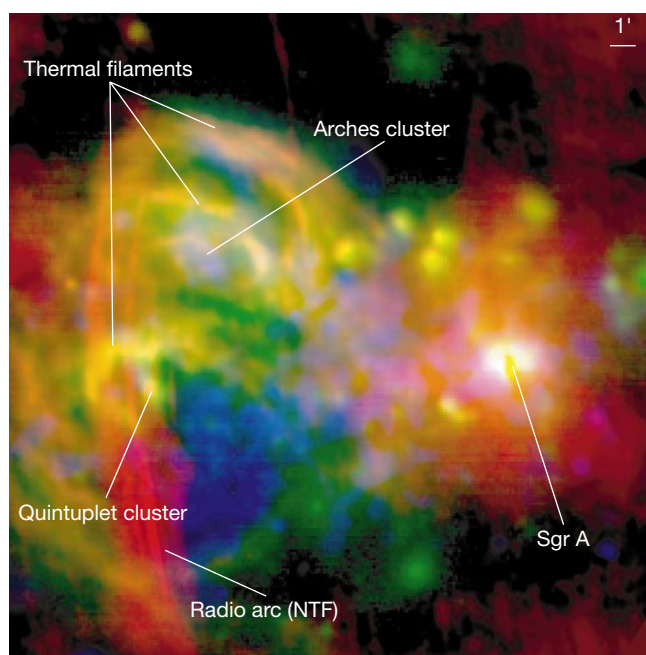


Figure 3 A multiwavelength close-up of the recent massive star-forming region near the Galactic centre. The colour image, plotted also in standard Galactic coordinates, is a composite of 20-cm radio continuum (red; ref. 26); 25- μ m mid-infrared (green; ref. 16); and 6.4-keV line emission (blue).

X-ray emission is globally correlated with massive star-forming regions, which include the Galactic Centre, Arches, and Quintuplet clusters²¹. In particular, the Quintuplet cluster is thought to be responsible for producing distinct shell-like structures as seen in radio and mid-infrared (Fig. 3). These structures, together with the enclosed diffuse X-ray-emitting materials, are probably a product of mechanical energy release from massive stars in the form of supernova explosions and fast stellar winds.

It is interesting to compare the massive star-forming regions of the Galactic Centre with the 30 Doradus nebula. The collective energy releases in the two regions are similar^{21,22}. With an overall extent of about 300 pc, the 30 Dor nebula consists of 'blisters' of X-ray-emitting gas enclosed in photon-ionized loops and shells²². But the bulk of the X-ray emission arises in gas with a characteristic temperature $\lesssim 10^7$ K, and the gas apparently cools off with increasing distance from the central cluster R136. However, such gas in the Galactic Centre region is not traced by X-rays observed in our Chandra survey, because of the heavy foreground absorption ($\sim 10^{23}$ cm⁻²). The strongly enhanced hard X-ray emission (≥ 4 keV) is thus unique to the Galactic Centre environment. The expansion and outflow of the high-pressure and buoyant plasma (or cosmic rays) may be responsible for the large-scale, diffuse radio and X-ray features which are oriented perpendicularly above and below the Galactic plane and are centred on the Galactic Centre region^{23,24}. □

Received 18 October; accepted 4 December 2001.

- Koyama, K. *et al.* ASCA view of our Galactic center: Remains of past activities in X-rays? *Publ. Astron. Soc. J.* **48**, 249–255 (1996).
- Sidoli, L. & Mereghetti, S. The X-ray diffuse emission from the Galactic center. *Astron. Astrophys.* **349**, L49–L52 (1999).
- Valinia, A. *et al.* On the origin of the iron K line in the spectrum of the Galactic X-ray background. *Astrophys. J.* **543**, 733–739 (2000).
- Tanaka, Y., Koyama, K., Maeda, Y. & Sonobe, T. Unusual properties of X-ray emission near the Galactic center. *Publ. Astron. Soc. J.* **52**, L25–L30 (2000).
- Wang, Q. D., Gotthelf, E. V., Lang, C. & Jagodzinski, F. UMass/Columbia Chandra X-ray survey of the Galactic center region. *Astrophys. J.* (submitted).
- Sidoli, L., Belloni, T. & Mereghetti, S. A catalogue of soft X-ray sources in the Galactic center region. *Astron. Astrophys.* **368**, 835–844 (2001).
- Sakano, M. *et al.* ASCA X-ray source catalogue in the Galactic center region. *Astrophys. J. Suppl. Ser.* (in the press); also as preprint astro-ph/0108376 at (<http://xxx.lanl.gov>) (2001).
- Ebisawa, K. *et al.* Origin of the hard X-ray emission from the Galactic plane. *Science* **293**, 1633–1635 (2001).
- Barret, D. *et al.* Hard X-ray emission from low-mass X-ray binaries. *Astrophys. J.* **533**, 329–351 (2000).
- Feng, Y. X. *et al.* Evolution of iron K α : Line emission in the black hole candidate GX339-4. *Astrophys. J.* **553**, 394–398 (2001).
- Yusef-Zadeh, F. *et al.* Detection of X-ray emission from the Arches cluster near the Galactic center. *Astrophys. J.* (in the press); also as preprint astro-ph/0108174 at (<http://xxx.lanl.gov>) (2001).
- Baganoff, F. *et al.* Rapid X-ray flaring from the direction of the supermassive black hole at the Galactic Centre. *Nature* **413**, 45–48 (2001).
- Murakami, H., Koyama, K. & Maeda, Y. Chandra observations of diffuse X-rays from the Sagittarius B2 cloud. *Astrophys. J.* **558**, 687–692 (2001).
- Maeda, Y. *et al.* A Chandra study of Sgr A East: A supernova remnant regulating the activity of our Galactic center? *Astrophys. J.* (in the press).
- Jackson, J. M., Heyer, M. H., Paglione, T. & Bolatto, A. HCN and CP in the central 630 parsecs of the Galaxy. *Astrophys. J.* **456**, L91–L95 (1996).
- Price, S. D. *et al.* Midcourse space experiment of the Galactic plane. *Astrophys. J.* **121**, 2819–2842 (2001).
- Frommerth, M. J., Melia, F. & Leahy, D. A. A Monte Carlo study of the 6.4 keV emission at the Galactic center. *Astrophys. J.* **547**, 129–132 (2001).
- Tsuboi, M., Ukita, N. & Handa, T. An expanding shell-like molecular cloud near the Galactic center arc. *Astrophys. J.* **481**, 263–266 (1997).
- Yusef-Zadeh, F., Wardle, M. & Parastaran, P. The nature of the Faraday screen toward the Galactic center nonthermal filament G359.54+0.18. *Astrophys. J.* **475**, 119–122 (1997).
- Boggs, S. E. *et al.* Diffuse galactic soft gamma ray emission. *Astrophys. J.* **544**, 320–329 (2000).
- Figer, D. *et al.* Hubble Space Telescope/NICMOS observations of massive stellar clusters near the Galactic center. *Astrophys. J.* **525**, 750–758 (1999).
- Wang, Q. D. Structure and evolution of hot gas in 30 Dor. *Astrophys. J.* **510**, L139–L143 (1999).
- Almy, R. C., McCammon, D., Digel, S. W., Bronfman, L. & May, J. Distance limits on the bright X-ray emission toward the Galactic center: Evidence for a very hot interstellar medium in the Galactic X-ray bulge. *Astrophys. J.* **545**, 290–300 (2000).
- Sofue, Y. Bipolar hypershell Galactic center starburst model: further evidence from ROSAT data and new radio and X-ray simulations. *Astrophys. J.* **540**, 224–235 (2000).
- Townsley, L. K., Broos, P. S., Garmire, G. & Nousek, J. Mitigating charge transfer inefficiency in the Chandra X-ray observatory advanced CCD imaging spectrometer. *Astrophys. J.* **534**, L139–L142 (2000).
- Yusef-Zadeh, F., Morris, M. & Chance, D. Large, highly organised radio structures near the galactic centre. *Nature* **310**, 557–561 (1984).

Acknowledgements

We thank F. Jagodzinski for assistance in the data calibration, and L. Townsley for helping with the CTI corrections.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Q.D.W. (e-mail: wqd@astro.umass.edu).

Transition-metal-based magnetic refrigerants for room-temperature applications

O. Tegus, E. Brück, K. H. J. Buschow & F. R. de Boer

Van der Waals-Zeeman Instituut, Universiteit van Amsterdam, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands

Magnetic refrigeration techniques based on the magnetocaloric effect (MCE) have recently been demonstrated as a promising alternative to conventional vapour-cycle refrigeration¹. In a material displaying the MCE, the alignment of randomly oriented magnetic moments by an external magnetic field results in heating. This heat can then be removed from the MCE material to the ambient atmosphere by heat transfer. If the magnetic field is subsequently turned off, the magnetic moments randomize again, which leads to cooling of the material below the ambient temperature. Here we report the discovery of a large magnetic entropy change in MnFeP_{0.45}As_{0.55}, a material that has a Curie temperature of about 300 K and which allows magnetic refrigeration at room temperature. The magnetic entropy changes reach values of

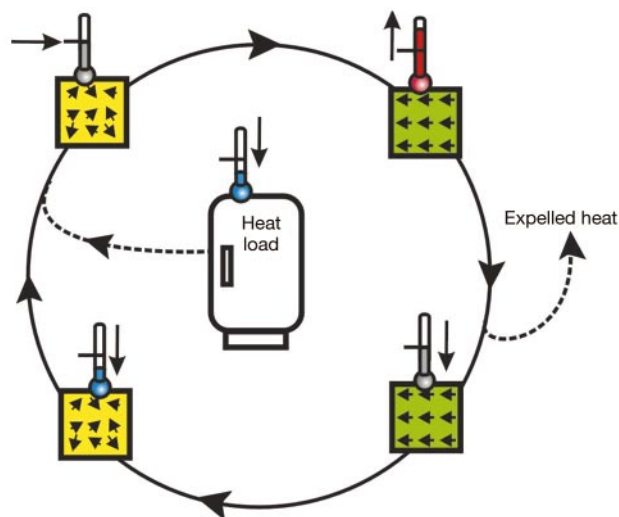


Figure 1 Schematic representation of a magnetic-refrigeration cycle, which transports heat from the heat load to its surroundings. Yellow and green depict the magnetic material in low and high magnetic fields, respectively. Initially randomly oriented magnetic moments are aligned by a magnetic field, resulting in heating of the magnetic material. This heat is removed from the material to its surroundings by a heat-transfer medium. On removing the field, the magnetic moments randomize, which leads to cooling of the magnetic material below the ambient temperature. Heat from the system to be cooled can then be extracted using a heat-transfer medium. Depending on the operating temperature, the heat-transfer medium may be water (with antifreeze) or air, and for very low temperatures, helium.

14.5 J K⁻¹ kg⁻¹ and 18 J K⁻¹ kg⁻¹ for field changes of 2 T and 5 T, respectively. The so-called giant-MCE material Gd₅Ge₂Si₂ (ref. 2) displays similar entropy changes, but can only be used below room temperature. The refrigerant capacity of our material is also significantly greater than that of Gd (ref. 3). The large entropy change is attributed to a field-induced first-order phase transition enhancing the effect of the applied magnetic field.

Magnetic refrigeration is an environmentally friendly cooling technology (see Fig. 1 for details). It does not use ozone-depleting chemicals (such as chlorofluorocarbons), hazardous chemicals (such as ammonia), or greenhouse gases (hydrochlorofluorocarbons and hydrofluorocarbons). Another important difference between vapour-cycle refrigerators and magnetic refrigerators is the amount of energy loss incurred during the refrigeration cycle. The cooling efficiency of magnetic refrigerators working with Gd has been shown⁴ to reach 60% of the theoretical limit, compared to only about 40% in the best gas-compression refrigerators. The use of magnetic refrigerators with such high energy efficiency will result in a reduced consumption of fossil fuels, in this way contributing to a reduced CO₂ release. However, with the currently available magnetic materials, this high efficiency is only realized in high magnetic fields of 5 T. We are therefore searching for new magnetic materials displaying larger MCEs, which then can be operated in lower fields of about 2 T that can be generated by permanent magnets.

The heating and cooling that occurs in the magnetic refrigeration technique is proportional to the size of the magnetic moments and to the applied magnetic field. This is why research in magnetic refrigeration is at present almost exclusively conducted on superparamagnetic materials and on rare-earth compounds⁵. For room-temperature applications like refrigerators and air-conditioners, compounds containing manganese should be a good alternative. Manganese is a transition metal of high abundance. Also, there exist (in contrast to rare-earth compounds) an almost unlimited number of manganese compounds with magnetic ordering temperatures near room temperature. However, the magnetic moment of manganese is generally only about half that of heavy rare-earth elements. Enhancement of the caloric effects associated with magnetic moment alignment may be achieved through the induction of a first-order phase transition, which will result in much higher efficiency of the magnetic refrigerator. In combination with currently available permanent magnets, this should open the way to the development of small-scale magnetic refrigerators that no longer rely on rather costly and service-intensive superconducting magnets. Another advantage of magnetocaloric refrigerators is that the cooling power can be varied by scaling from milliwatts to a few hundred watts.

Following the discovery of a sub-room-temperature giant MCE

in the ternary compound Gd₅(Ge_{1-x}Si_x)₄ (0.2 ≤ x ≤ 0.5), there is incentive from both fundamental and practical points of view to study the MCE in Gd-based materials⁶⁻⁸. The common feature of these compounds is that they undergo a first-order structural and magnetic phase transition, which leads to a giant magnetic-field-induced entropy change, across their ordering temperature. Many manganese compounds, such as MnAs (ref. 9), have a variety of structural and magnetic phase transitions. For a working material in a magnetic refrigerator it is, however, essential that the phase transition is also reversible. Here we report on MnFeP_{1-x}As_x compounds exhibiting a large magnetic entropy change, of similar magnitude to that in the so-called giant MCE material Gd₅Ge₂Si₂, but whose optimum temperature of use is at and above room temperature. This result is of practical importance, because it shows that these compounds could be good working materials for magnetic refrigeration in household refrigerators or air-conditioning.

In the intermediate composition range (0.15 ≤ x ≤ 0.66), compounds in the MnFeP_{1-x}As_x system crystallize in the hexagonal Fe₂P-type of structure. They have interesting magnetic properties, including a field-induced first-order magnetic transition¹⁰. For studying the room-temperature MCE in this system, we synthesized polycrystalline samples. The binary compounds Fe₂P and FeAs₂ were mixed in the appropriate proportions with Mn chips and P powder, and ground by ball milling under a protective atmosphere. The powders obtained were sealed in molybdenum tubes under an argon atmosphere, heated at 1,273 K for 120 h, followed by a homogenization process at 923 K for 120 h and finally by slow cooling to ambient conditions. The powder X-ray diffraction patterns show that the compound crystallizes in the hexagonal Fe₂P-type structure. In this structure, the Mn atoms occupy the 3(g) sites, the Fe atoms occupy the 3(f) sites and the P and the As atoms occupy 2(c) and 1(b) sites statistically¹¹. From the broadening of the X-ray diffraction reflections, we estimate the average grain size to be about 100 nm.

Figure 2 shows the temperature dependence of the magnetization near room temperature of MnFeP_{0.45}As_{0.55} and Gd (Alfa Aesar 3N), as determined in an applied magnetic field of 1 T. The change in magnetization in MnFeP_{0.45}As_{0.55} at the ordering temperature T_C is much larger than that of Gd, despite the fact that the magnetic moment of Gd is much larger at low temperatures (248 A m² kg⁻¹ compared to 125 A m² kg⁻¹). Variation of the P/As ratio between 3/2 and 1/2 makes it possible to tune T_C and the optimal operating temperature between 200 and 350 K (-70 °C and 80 °C), without losing the large MCE. Figure 2 shows magnetization measured with increasing (heating) and decreasing (cooling) temperature. The thermal hysteresis is a signature of a first-order phase transition. Because of the small size of the thermal hysteresis (less than 1 K), the

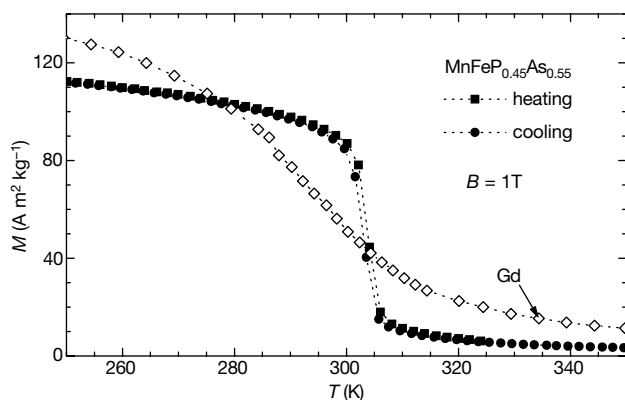


Figure 2 Temperature dependence of the magnetization M for MnFeP_{0.45}As_{0.55} and Gd. Data are shown for increasing (heating) and decreasing (cooling) temperature in a field of 1 T.

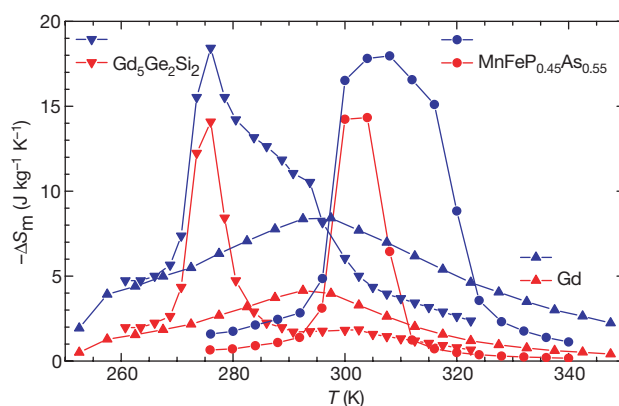


Figure 3 Magnetic entropy changes of MnFeP_{0.45}As_{0.55}, Gd and Gd₅Ge₂Si₂. Data are shown for external magnetic field changes of 0 to 2 T (red), and 0 to 5 T (blue), calculated from the magnetization data. Circles, MnFeP_{0.45}As_{0.55}; triangles, Gd; and inverted triangles, Gd₅Ge₂Si₂ (after ref. 2).

magnetization process can be considered as being reversible in temperature. From the magnetization curve at 5 K, the saturation magnetization was determined as $3.9 \mu_B$ per formula unit. This high magnetization originates from the parallel alignment of the Mn and Fe moments, though the moments of Mn are much larger than those of Fe (ref. 12). Variation of the Mn/Fe ratio may also be used to further improve the MCE.

The magnetic entropy changes, ΔS_m , are calculated from magnetization data by means of the equation

$$\Delta S_m(T, B) = S_m(T, B) - S_m(T, 0) = \int_0^B \left(\frac{\partial M}{\partial T} \right) dB'$$

which is based on a Maxwell relation. The results are shown in Fig. 3. The calculated maximum values of the magnetic entropy changes are $14.5 \text{ J kg}^{-1} \text{ K}^{-1}$ and $18 \text{ J kg}^{-1} \text{ K}^{-1}$ for field changes from 0 to 2 T and 0 to 5 T, respectively. The maximum magnetic entropy in 3d materials depends on the spin moment S . Because there are two magnetic ions per formula unit, we have $S_m = 2R \ln(2S + 1)$, where R is the universal gas constant. From the saturation magnetic moment, we estimate the average S value of the magnetic ions to be 1, thus $S_m = 18.3 \text{ J mol}^{-1} \text{ K}^{-1} = 117 \text{ J kg}^{-1} \text{ K}^{-1}$, which is about 6 times larger than the value obtained from the magnetization measurements. For comparison, the magnetic entropy change of the compound $\text{Gd}_5\text{Ge}_2\text{Si}_2$ and Gd is also shown in Fig. 3. It is evident that the MCE in $\text{MnFeP}_{1-x}\text{As}_x$ compounds is comparable with that of $\text{Gd}_5\text{Ge}_2\text{Si}_2$, though the Gd compound has a larger magnetic moment at 5 K. The origin of the large magnetic entropy change is in the comparatively high 3d moments and in the rapid change of magnetization in the field-induced magnetic phase transition. In rare-earth materials, the magnetic moment fully develops only at low temperatures, and therefore the entropy change near room temperature is only a fraction of the potential value. In 3d compounds, the strong magneto-crystalline coupling results in competing intra- and inter-atomic interactions, and leads to a modification of metal-metal distances which may change the iron and manganese magnetic moments and favour spin ordering.

The excellent magnetocaloric features of the compound $\text{MnFeP}_{0.45}\text{As}_{0.55}$, in addition to the very low material costs, make it an attractive candidate material for a commercial magnetic refrigerator.

Received 30 July; accepted 6 November 2001.

- Glanz, J. Making a bigger chill with magnets. *Science* **279**, 2045 (1998).
- Pecharsky, V. K. & Gschneidner, K. A. Jr Giant magnetocaloric effect in $\text{Gd}_5(\text{Si}_2\text{Ge}_2)$. *Phys. Rev. Lett.* **78**, 4494–4497 (1997).
- Gschneidner, K. A. Jr *et al.* Recent developments in magnetic refrigeration. *Mater. Sci. Forum* **315–317**, 69–76 (1999).
- Zimm, C. *et al.* Description and performance of a near-room temperature magnetic refrigerator. *Adv. Cryogen. Eng.* **43**, 1759–1766 (1998).
- Tishin, A. M. in *Handbook of Magnetic Materials* Vol. 12 (ed. Buschow, K. H. J.) 395–524 (North Holland, Amsterdam, 1999).
- Choe, W. *et al.* Making and breaking covalent bonds across the magnetic transition in the giant magnetocaloric material $\text{Gd}_5(\text{Si}_2\text{Ge}_2)$. *Phys. Rev. Lett.* **84**, 4617–4620 (2000).
- Giguere, A. *et al.* Direct measurement of the “giant” adiabatic temperature change in $\text{Gd}_5(\text{Si}_2\text{Ge}_2)$. *Phys. Rev. Lett.* **83**, 2262–2265 (1999).
- Morellon, L. *et al.* Nature of the first-order antiferromagnetic-ferromagnetic transition in the Ge-rich magnetocaloric compounds $\text{Gd}_5(\text{Si}_x\text{Ge}_{1-x})_4$. *Phys. Rev. B* **62**, 1022–1026 (2000).
- Pytlík, L. & Zieba, A. Magnetic phase diagram of MnAs . *J. Magn. Magn. Mater.* **51**, 199–210 (1985).
- Zach, R., Guillot, M. & Fruchart, R. The influence of high magnetic fields on the first order magneto-elastic transition in $\text{MnFe}(\text{P}_{1-x}\text{As}_x)$ systems. *J. Magn. Magn. Mater.* **89**, 221–228 (1990).
- Bacmann, M. *et al.* Magnetoelastic transition and antiferro-ferromagnetic ordering in the system $\text{MnFeP}_{1-x}\text{As}_x$. *J. Magn. Magn. Mater.* **134**, 59–67 (1994).
- Beckman, O. & Lundgren, L. in *Handbook of Magnetic Materials* (ed. Buschow, K. H. J.) Vol. 6, 181–287 (North Holland, Amsterdam, 1991).

Acknowledgements

We thank A.J. Riemersma for preparation of graphs. Part of this work was performed within the scientific exchange program between the Netherlands and China. This work was financially supported by the Dutch Technology Foundation STW.

Correspondence and requests for materials should be addressed to E.B. (e-mail: bruck@wins.uva.nl).

Remote electronic control of DNA hybridization through inductive coupling to an attached metal nanocrystal antenna

Kimberly Hamad-Schifferli*, John J. Schwartz†, Aaron T. Santos*, Shuguang Zhang‡ & Joseph M. Jacobson*

* The Media Laboratory and the ‡Center for Biomedical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, Massachusetts 02139, USA

† Engeneos, 40 Bear Hill Road, Waltham, Massachusetts 02451, USA

Increasingly detailed structural¹ and dynamic^{2,3} studies are highlighting the precision with which biomolecules execute often complex tasks at the molecular scale. The efficiency and versatility of these processes have inspired many attempts to mimic or harness them. To date, biomolecules have been used to perform computational operations⁴ and actuation⁵, to construct artificial transcriptional loops that behave like simple circuit elements^{6,7} and to direct the assembly of nanocrystals⁸. Further development of these approaches requires new tools for the physical and chemical manipulation of biological systems. Biomolecular activity has been triggered optically through the use of chromophores^{9–14}, but direct electronic control over biomolecular ‘machinery’ in a specific and fully reversible manner has not yet been achieved. Here we demonstrate remote electronic control over the hybridization behaviour of DNA molecules, by inductive coupling of a radio-frequency magnetic field to a metal nanocrystal covalently linked to DNA¹⁵. Inductive coupling to the nanocrystal increases the local temperature of the bound DNA, thereby inducing denaturation while leaving surrounding molecules relatively unaffected. Moreover, because dissolved biomolecules dissipate heat in less than 50 picoseconds (ref. 16), the switching is fully reversible. Inductive heating of macroscopic samples is widely used^{17–19}, but the present approach should allow extension of this concept to the control of hybridization and thus of a broad range of biological functions on the molecular scale.

Inductive coupling is the transfer of energy between circuits. If the secondary circuit has finite impedance, eddy currents are produced which are converted to heat by the Joule effect¹⁷. Heating a conductor by placing it in an alternating magnetic field is generally used to heat macroscopic samples. Here we apply it to metallic nanocrystals (diameter 1.4 nm) in solution. Induction heating is accompanied by a skin-depth effect, resulting from partial cancellation of the magnetic fields. Consequently, most of the power absorbed by the conductor is concentrated in a depth d_0 , given by:

$$d_0 = \frac{1}{2\pi} \sqrt{\frac{\rho 10^7}{\mu_r \mu_0 f}} \quad (1)$$

where μ_r is magnetic permeability, μ_0 is the permeability of free space, ρ is the material resistivity, and f is the frequency of the alternating magnetic field. The power density P is given by:

$$P = 4\pi H_c^2 \mu_0 \mu_r f \frac{d_0}{d} \quad (2)$$

where d is the sample diameter, H_c is the magnetic field strength, and F is a transmission factor that has a sigmoidal dependence on d/d_0 . Optimal power absorption and heating occur when $d/d_0 = 3.5$. To optimally heat by induction a gold nanocrystal with $d = 1.4 \text{ nm}$, $f = 49 \text{ GHz}$ (radio frequency range) is required^{17,18}. Here we use

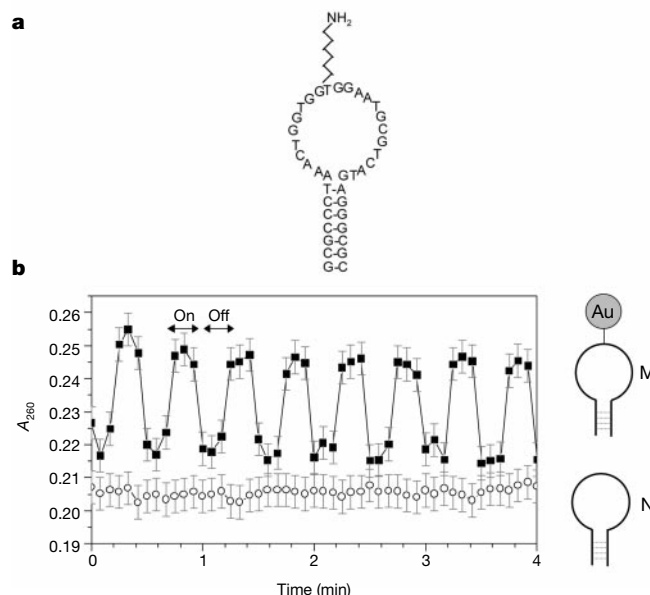


Figure 1 Inductive coupling to nanocrystals linked to DNA and evidence of dehybridization. **a**, Sequence of the hairpin molecule, M. The molecule is self-complementary at the ends for 7 bases, with a primary amine in the loop to which a 1.4-nm gold nanocrystal is covalently linked. **b**, Absorbance at 260 nm (A_{260}) of a solution of M in RFMF (squares). Arrows indicate when the RFMF is on/off. Circles, response of N (no nanocrystals) in RFMF. Gold nanocrystals also show an absorbance increase under

RFMF, qualitatively attributed to the change in the optical absorption due to eddy currents. However, the absorbance increases nearly uniformly over 200–300 nm with RFMF and is small relative to the oligo change in absorbance. To obtain the data shown here, the nanocrystal contribution was subtracted (see Supplementary Information for details). M with a fluorophore/quencher pair on the ends shows fluorescence increasing reversibly with RFMF (not shown), confirming dehybridization.

$f = 1$ GHz, which results in a skin depth larger than the particle so that the entire particle is heated. We note that additional mechanisms, including induced dipole torque on the nanoparticle which increases the kinetic energy of the nanoparticles, may be important at these length scales. The physics of heating nanometre particles has not been explored previously, and will be the subject of future investigations. There are numerous examples of the heating of magnetic particles by a magnetic field for the treatment of tumours through macroscopic tissue heating¹⁹. However, we emphasize that the (magnetite) particles in such cases are several orders of magnitude larger in volume than the particles used in our investigation, and are not interfaced to individual molecules.

In order to demonstrate reversible electronic control, we have constructed a DNA hairpin-loop oligonucleotide covalently linked to a nanometre scale antenna. 38-nucleotide hairpin-loop DNA which is self-complementary at each end for 7 bases²⁰ (Fig. 1a) was covalently linked to a 1.4-nm gold nanocrystal by a primary amine appended to a single base^{21–24}. Because of the loop constraint, the DNA rehybridizes on a timescale comparable to dehybridization²⁵. The dehybridization was monitored by the hyperchromicity of DNA, measured by optical absorbance at 260 nm (A_{260})²⁶. A solution of the nanocrystal-linked oligonucleotides (referred to as M) was put into a radio-frequency magnetic field (RFMF) with $f = 1$ GHz. A_{260} was monitored in an ultraviolet–visible spectro-

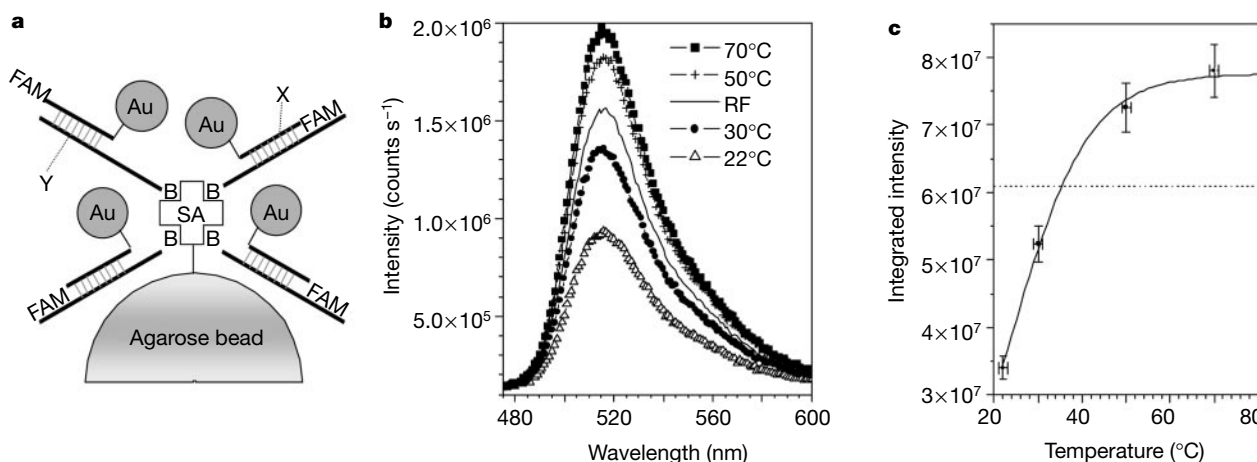


Figure 2 Determination of effective temperature from inductive coupling to a gold nanocrystal linked to DNA. **a**, The two-phase system. X is a 12-nucleotide DNA molecule covalently labelled with a gold nanocrystal on its 3' end and the fluorophore FAM on its 5' end. Its complementary strand Y has biotin (B) on the 5' end, which was captured onto agarose beads with streptavidin (SA) on the surface. Once X dehybridizes from Y it can

diffuse into the supernatant. **b**, Supernatant fluorescence spectra. Each sample is subject to a fixed temperature (22°C, 30°C, 50°C, 70°C) or RFMF (solid line). **c**, Integrated peak intensity of the supernatant fluorescence spectra shown in **b** (circles) and a sigmoidal fit (solid line), and the intensity of the sample exposed to RFMF (dotted line). Extrapolation of RFMF sample intensity results in an effective temperature of 35°C for X.

photometer as the RFMF was pulsed at 15-s intervals (Fig. 1b). As the RFMF was switched on, A_{260} increased from 0.22 to 0.25, indicating that M dehybridized with the RFMF (squares in Fig. 1b). When the RFMF was switched off, A_{260} returned to its original value. Cycling through the on and off states was repeatable. The on/off ratio of A_{260} is 1.09, consistent with calculated absorption coefficients for this sequence hybridizing and dehybridizing 7 bases. Control solutions with DNA only (that is, not linked to gold nanocrystals), referred to as N, resulted in no change in the absorbance with RFMF (circles in Fig. 1b). This shows that inductive coupling to a covalently bound metal nanocrystal can reversibly dehybridize DNA on a timescale of at most several seconds.

To determine the effective temperature that inductive coupling to the nanocrystal produces locally, we designed a two-phase system in which DNA is dehybridized from a solid support into solution (Fig. 2a). X is a 12-nucleotide DNA molecule with a nanocrystal on

the 3' end and a fluorophore (FAM) on the 5' end (see Methods). It was hybridized to its complement Y, which was immobilized onto 100- μm streptavidin-coated agarose beads comprising the solid phase. As X dehybridizes from Y, it diffuses into the supernatant. The amount of dehybridization is proportional to the amount of X in the supernatant, which was measured by fluorescence spectroscopy. One aliquot of the sample was exposed to the RFMF and its supernatant removed. Its fluorescence spectrum was compared to supernatants of aliquots thermally heated at specific temperatures. The intensity of the supernatant fluorescence spectra (Fig. 2b) increases with temperature, indicative of increased amounts of X dehybridized. The sample exposed to RFMF has intensity (solid line) in between the 30 °C and 50 °C samples. The integrated peak intensity of the spectra (Fig. 2c) as a function of temperature (squares) can be fitted to a sigmoidal (solid line), corresponding to a thermal denaturation curve. The intensity of the RFMF sample (dashed line) extrapolates to 35 °C, indicating that the effective temperature increase on X produced by the RFMF is +13 °C. This change in temperature is sufficient for control of many biological processes.

One important property is the ability to address molecules with a nanocrystal while having a lesser effect on molecules not bound to a nanocrystal. If induction heating of the nanocrystal is sufficiently spatially localized, it will afford selectivity. We used a two-phase system in which X was mixed with Z, identical in sequence to X but instead has TMR (see Methods) on the 3' end (Fig. 3a). TMR emits at a wavelength distinct from FAM ($\lambda_{\text{max}} = 563 \text{ nm}$). The two-phase system comprises tetrameric avidin acrylic beads with both X-Y and Z-Y hybrids on the surface in approximately equimolar amounts. One sample was exposed to RFMF, and the supernatant was compared to a thermally heated sample (70 °C) by fluorescence. The difference of the spectrum before and after dehybridization for the 70 °C sample (Fig. 3b, right) has peaks at 515 nm and 563 nm, indicating dehybridization of both X and Z. The RFMF sample (Fig. 3b, left) shows intensity at 515 nm due to dehybridized X but negligible intensity at 563 nm, which indicates selectivity: Z is relatively unaffected by induction heating of X in its proximity. Quantification of the spectra yields ~80% X for the RFMF sample and ~55% for the 70 °C sample. The slight amount of Z dehybridized in the RFMF sample is presumably due to the proximity of the molecules associated with tetrameric avidin (intermolecular separation is expected to be $\geq 10 \text{ nm}$).

RFMF experiments on solutions of mixtures of M and N also illustrate selective heating. Figure 3c shows A_{260} with RFMF on and off for a fixed concentration of M as N was added. Values of A_{260} for the on and off states increase linearly, but the difference (ΔA_{260}) remains constant. Samples of increasing concentration of M show ΔA_{260} increasing with concentration (not shown). These experiments indicate that induction heating of the nanocrystal on M is sufficiently localized that N is unaffected. Another indication of selectivity of RFMF dehybridization is evident in the purification of gold-linked X from unlabelled X, where the fraction of gold-linked oligonucleotides increases with successive RFMF dehybridizations. Figure 3d shows the fluorescence spectra of a sample before and after RFMF purification. A sample of partially labelled X is RFMF dehybridized and compared to a 70 °C aliquot, which effectively provides the total number of X in the system (dots). If the RFMF supernatant is rehybridized to Y and RFMF dehybridized a second time, its intensity is higher relative to the 70 °C aliquot (lines). The percentage of molecules dehybridized by RFMF increases after purification, suggesting the RFMF preferentially dehybridizes X linked with nanocrystals. If RFMF dehybridization were non-selective, this percentage would be the same for both samples. These experiments show that induction heating of the nanocrystal is localized such that surrounding molecules are not substantially affected.

Manipulation of DNA by itself is useful, as it has potential as an

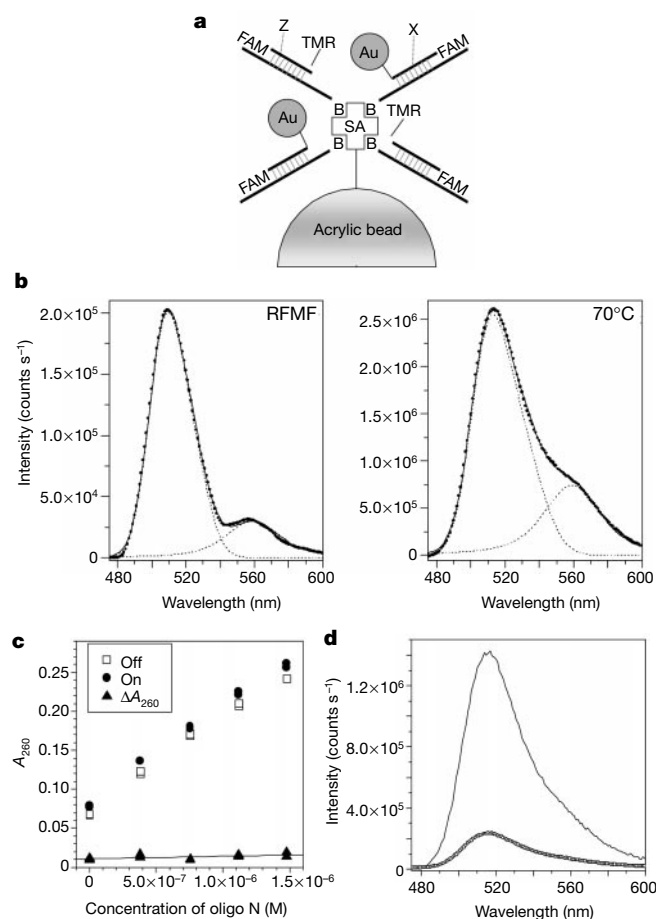


Figure 3 Testing selectivity. **a**, Two-phase system. Z is identical to X but has the fluorophore TMR on the 3' end ($\lambda_{\text{max}} = 563 \text{ nm}$). The beads have both X-Y and Z-Y hybrids. **b**, Difference fluorescence spectra (dots) (after dehybridization by heat/RFMF minus before dehybridization) scaled to same height. Shown are individual fluorescence peak fits (dashed lines) and composite fits (solid lines). When fitting the experimental data obtained upon RFMF and heat exposure, two spectral peaks centred at 516 nm and 563 nm and having full-widths at half-maximum of 31 nm and 38 nm, respectively, were used in each case. Sample exposed to RFMF (left) and to 70 °C (right). **c**, A_{260} of M and N mixture with RFMF on (circles) and off (squares), ΔA_{260} (triangles), and fit (line) as a function of N concentration. The concentration of M is $1.12 \times 10^{-7} \text{ M}$ while N ranges from 0 to $1.4 \times 10^{-6} \text{ M}$. Based on concentrations the estimated density is ~ 1 molecule per 10^6 nm^3 , an upper limit on intermolecular separation as it does not account for diffusion. **d**, Spectra of supernatants with one (dots) or two (line) RFMF dehybridizations normalized to the intensity of the 70 °C aliquot for each sample.

actuator⁵ and performing computations²⁷. Because nanocrystals can be readily attached to proteins, switching more complex processes—such as enzymatic activity (S.Z., J. Shi, M. Jura, K.H.-S., J.J.S. and J.M.J., manuscript in preparation), biomolecular assembly, gene regulation, and protein function—might be possible. Due to the spatial localization of denaturation, it might be possible to control portions of proteins or nucleic acids while the rest of the molecule and neighbouring species would remain relatively unaffected. Control of dehybridization of antisense oligos in transcription²⁸ would permit reversible control of the production of specific messenger RNA. Because the addressing is not optical, this technology could be useful in highly scattering media. □

Methods

Linking gold nanocrystals to DNA hairpin-loop molecules

DNA oligonucleotides were synthesized via solid-phase synthesis (Research Genetics) with addition of internal amines, terminal biotin, or fluorophores. DNA hairpin-loop oligonucleotides of the type used to make M ('molecular beacons'²⁰) have a thymine in the loop region which was modified on the 5' position of the base with a C6 primary amine, to which a nanocrystal was covalently linked. 1.4-nm uncharged gold nanocrystals functionalized with a single sulpho *N*-hydroxy-succinimide (NHS) ester²⁹ (Nanoprobes) were incubated with the oligonucleotide in the manufacturer's buffer at room temperature for 1 h, resulting in a covalent bond between the nanocrystal and oligonucleotide³⁰. Nanocrystals were in molar excess (>10 ×). Excess nanocrystals were removed by 70% ethanol precipitation on ice and repeated washes. The nanocrystals are soluble in ethanol, but the DNA or species attached to it precipitates. Gel electrophoresis is effective in separating nanocrystals from DNA-labelled gold nanocrystals, but for short DNA the mobility shift is expected to be too small to resolve the two species²⁴. The concentration of M and N is 1×10^{-6} M in 1X PBS (1X PBS = 1 mM KH_2PO_4 , 10 mM Na_2HPO_4 , 137 mM NaCl, 2.7 mM KCl).

RFMF

Alternating magnetic fields were generated by applying an alternating current to a coil with 35 turns and a cross section of $\sim 1 \text{ cm}^2$. Coils were wrapped around a plastic cuvette/tube holder with open structures to maximize light passage. Currents with $f = 1 \text{ GHz}$ were obtained using an RF signal generator (Hewlett Packard 8648C) with an output of 1 mW in conjunction with a linear amplifier. The ultimate output power range used was 0.4–4 W, though these are upper limits to the exact power inside the coil due to losses from set-up architecture (10% input power estimated to be transferred to coil). DNA hairpin samples (volume 200 μl) were in a 3 mm × 3 mm quartz cuvette inside the coil. The two-phase samples were in 250- μl plastic tubes (supernatant volume 165 μl , solid phase volume 60 μl) put into the coil in a water bath. The power used for the experiments shown in Figs 1 and 2 was 4 W, and 1 W for those in Fig. 3.

Spectroscopy

UV–visible spectra were taken on a DU530 Beckman spectrophotometer with time resolution limited to 5–s steps. Fluorescence spectra was taken on a Spex Fluoromax fluorometer with 1-nm steps, 0.1-s integration time, entrance/exit slits 1–5 nm, $\lambda_{\text{excitation}} = 450 \text{ nm}$. Signal was averaged multiple times (≤ 10). Fluorescence peak areas were quantified by Voigt lineshape deconvolution using a commercial program (Peakfit, SPSS). Figure 3b fits had the same parameters for peak positions and widths for both spectra.

Two-phase system

Molecule X is a 12-nucleotide DNA that has a 6-carboxyfluorescein (FAM, $\lambda_{\text{max}} = 516 \text{ nm}$) on the 5' end and a C6 primary amine on the 3' end ($T_m = 40^\circ\text{C}$). It was labelled with the nanocrystal as described above. Molecule Y is a 68-nucleotide DNA complementary to X, and has a biotin on the 5' end. X and Y were hybridized, and the X–Y hybrid was incubated with streptavidin agarose beads (Sigma-Aldrich) with $\langle d \rangle = 100 \mu\text{m}$ to allow the biotin on Y to bind to streptavidin. Samples were washed several times with 1X PBS to remove free gold. This biotin–streptavidin reaction ($K_d = 10^{-15} \text{ M}$) is robust for the temperatures used here³⁰. Tests in which bead-bound Y was heated to 85°C showed no presence of oligonucleotide in the supernatant, confirming the biotin–streptavidin interaction was intact.

In the construction of the two-phase system (Fig. 2), X was purified from unlabelled oligonucleotides by exposing bead-immobilized X–Y to RFMF. Oligonucleotides not linked to a nanocrystal had no response to RFMF and thus remain hybridized to Y. This is another alternative to purification by gel electrophoresis²⁴. Purified X was hybridized to Y, then exposed to streptavidin agarose beads to immobilize X–Y. After multiple washes with 1X PBS, buffer was added to the beads to serve as the liquid phase. Samples were exposed to RFMF or heated conventionally at a specific temperature for 5 min. Following heating or exposure to the RFMF, the supernatants were separated from the beads and then measured by fluorescence spectroscopy. All fluorescence spectra were taken under identical conditions (temperature, spectrometer parameters).

Mixed two-phase system

Oligonucleotide Z was identical in sequence to X but was labelled on the 3' end with

NHS-tetramethylrhodamine (TMR, Molecular Probes). Attachment of the TMR to the primary amine on Z was achieved by incubating the molecules together in sodium bicarbonate solutions (1 h, room temperature) in the dark to prevent photobleaching (manufacturer directions). Oligonucleotide Z was hybridized to Y, mixed in solution with X–Y, then incubated with tetrameric avidin acrylic beads (Sigma-Aldrich) as described above. Free molecules were removed with multiple washes of 1X PBS. The sample is equimolar in X to Z on the beads. The surface density of the monomeric avidin on the beads is ~ 1 molecule per nm^2 .

Received 12 July; accepted 16 November 2001.

- Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
- Deniz, A. A. *et al.* Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc. Natl Acad. Sci. USA* **97**, 5179–5184 (2000).
- Davenport, R., Wuite, G., Landick, R. & Bustamante, C. Single-molecule study of transcriptional pausing and arrest by *E. coli* RNA polymerase. *Science* **287**, 2497–2500 (2000).
- Winfree, E., Liu, F., Wenzler, L. & Seeman, N. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539–44 (1998).
- Yurke, B. *et al.* A DNA-fuelled molecular machine made of DNA. *Nature* **406**, 605–608 (2000).
- McAdams, H. H. & Arkin, A. Gene regulation: towards a circuit engineering discipline. *Curr. Biol.* **10**, R318–R320 (2000).
- Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
- Whaley, S. R., English, D. S., Hu, E. L., Barbara, P. F. & Belcher, A. M. Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly. *Nature* **405**, 665–668 (2000).
- Telford, J. R., Wittung-Stafshede, P., Gray, H. B. & Winkler, J. R. Protein folding triggered by electron transfer. *Acc. Chem. Res.* **31**, 755–763 (1998).
- Monroe, W. T., McQuain, M. M., Chang, M. S., Alexander, J. S. & Haselton, F. R. Targeting expression with light using caged DNA. *J. Biol. Chem.* **274**, 20895–20900 (1999).
- Asanuma, H., Yoshida, T., Ito, T. & Komiyama, M. Photo-responsive oligonucleotides carrying azobenzene at the 2'-position of uridine. *Tetrahedr. Lett.* **40**, 7995–7998 (1999).
- Liu, D., Karanicolas, J., Yu, C., Zhang, Z. & Woolley, G. A. Site-specific incorporation of photo-isomerizable azobenzene groups into Ribonuclease S. *Bioorg. Med. Chem. Lett.* **7**, 2677–2680 (1997).
- Haupts, U., Tittor, J. & Oesterhelt, D. Closing in on bacteriorhodopsin: progress in understanding the molecule. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 367–99 (1999).
- Ashkenazi, G., Ripoll, D. R., Lotan, N. & Scheraga, H. A. A molecular switch for biochemical logic gates: Conformational studies. *Biosens. Bioelectron.* **12**, 85–95 (1997).
- Hamad-Schiffertli, K., Schwartz, J. J., Santos, A. T., Zhang, S. & Jacobson, J. M. in *Mater. Res. Soc. Proc.* (eds Hahn, H. W., Feldheim, D. L., Kubiak, C. P., Tannenbaum, R. & Siegel, R. W.) Y8.43 (MRS, San Francisco, 2001).
- Lian, T., Locke, B., Kholodenko, Y. & Hochstrasser, R. M. Energy flow solute to solvent probed by femtosecond IR spectroscopy: malachite green and heme protein solutions. *J. Phys. Chem.* **98**, 11648–11656 (1994).
- Orfeuil, M. *Electric Process Heating: Technologies/ Equipment/ Applications* (Battelle Press, Columbus, Ohio, 1987).
- Zinn, S. & Semiatin, S. L. *Elements of Induction Heating, Design Control, and Applications* (ASM International, Materials Park, Ohio, 1988).
- Hergt, R. *et al.* Physical limits of hyperthermia using magnetite fine particles. *IEEE Trans. Magn.* **34**, 3745–3754 (1998).
- Bonnet, G., Tyagi, S., Libchaber, A. & Kramer, F. R. Thermodynamic basis of the enhanced specificity of structured DNA probes. *Proc. Natl Acad. Sci. USA* **96**, 6171–6176 (1999).
- Taton, A. T., Mirkin, C. A. & Letsinger, R. L. Scanometric DNA array detection with nanoparticle probes. *Science* **289**, 1757–1760 (2000).
- Loweth, C. J., Caldwell, W. B., Peng, X., Alivisatos, A. P. & Schultz, P. G. DNA-based assembly of gold nanocrystals. *Angew. Chem. Int. Edn Engl.* **38**, 1808–1812 (1999).
- Mattoussi, H. *et al.* Self-assembly of CdSe–ZnS quantum dot bioconjugates using an engineered recombinant protein. *J. Am. Chem. Soc.* **122**, 12142–12150 (2000).
- Zanchet, D., Michel, C. M., Parak, W. J., Gerion, D. & Alivisatos, A. P. Electrophoretic isolation of discrete Au nanocrystal/DNA conjugates. *Nanotechnology* **1**, 32–35 (2001).
- Bonnet, G., Krichavsky, O. & Libchaber, A. Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc. Natl Acad. Sci. USA* **95**, 8602–8606 (1998).
- Cantor, C. R. & Schimmel, P. *Biophysical Chemistry* (Freeman, San Francisco, 1980).
- Mao, C., LaBean, T. H., Reif, J. H. & Seeman, N. C. Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. *Nature* **407**, 493–496 (2000).
- Probst, J. C. Antisense oligodeoxynucleotide and ribozyme design. *Methods* **22**, 271–281 (2000).
- Reardon, J. E. & Frey, P. A. Synthesis of undecagold cluster molecules as biochemical labeling reagents. 1. Monoacyl and mono[N-succinimidooxy]undecagold clusters. *Biochemistry* **23**, 3849–3856 (1984).
- Hermanson, G. T. *Bioconjugate Techniques* (Academic, San Diego, 1996).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank N. Afeyan, E. Lander and P. Matsudaira for discussions in the early stages of this work, and J.P. Shi for comments on the experimental work. This work was supported by DARPA and the MIT Media Lab TTT consortium.

Correspondence and requests for materials should be addressed to J.M.J. (e-mail: jacobson@media.mit.edu) or S.Z. (e-mail: shuguang@mit.edu).

Reduced nitrogen fixation in the glacial ocean inferred from changes in marine nitrogen and phosphorus inventories

Raja S. Ganeshram*, Thomas F. Pedersen†, Stephen E. Calvert‡ & Roger François‡

*Department of Geology and Geophysics, University of Edinburgh, West Mains Road, Edinburgh EH9 3JW, UK

†Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, V6T 1Z4, Canada

‡Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02540, USA

To explain the lower atmospheric CO₂ concentrations during glacial periods, it has been suggested that the productivity of marine phytoplankton was stimulated by an increased flux of iron-bearing dust to the oceans^{1,2}. One component of this theory is that iron—an essential element/nutrient for nitrogen-fixing organisms—will increase the rate of marine nitrogen fixation, fuelling the growth of other marine phytoplankton and increasing CO₂ uptake. Here we present data that questions this hypothesis. From a sediment core off the northwestern continental margin of Mexico, we show that denitrification and phosphorite formation—processes that occur in oxygen-deficient upwelling regions, removing respectively nitrogen and phosphorus from the ocean—declined in glacial periods, thus increasing marine inventories of nitrogen and phosphorus. But increases in phosphorus were smaller and less rapid, leading to increased N/P ratios in the oceans. Acknowledging that phytoplankton require nitrogen and phosphorus in constant proportions, the Redfield ratio³, and that N/P ratios greater than the Redfield ratio are likely to suppress nitrogen fixation^{4,5}, we suggest therefore that marine productivity did not increase in glacial periods in response to either increased nutrient inventories or greater iron supply.

Denitrification of fixed nitrogen represents a substantial loss of biologically available nitrogen from the ocean. It occurs mostly in oxygen-deficient zones where large settling fluxes of organic detritus in coastal upwelling areas near continental margins leads to consumption of oxygen in subsurface waters at rates that exceed resupply from horizontal and vertical mixing. This leads to water-column denitrification because, in the absence of oxygen, NO₃⁻ is used as an electron acceptor in the bacterially mediated degradation of organic matter. Because the gaseous products of denitrification (N₂O and N₂) are to a large extent lost to the atmosphere, this process constitutes a net loss of fixed nitrogen from the ocean^{6,7}. Water-column denitrification occurs today primarily in three areas: off northwestern Mexico in the eastern tropical North Pacific Ocean, off Peru–Chile in the eastern tropical South Pacific Ocean and in the Arabian Sea⁸. The loss of fixed N in these three regions accounts for almost one-half of the total oceanic nitrogen loss of 200 Tg N yr⁻¹ (ref. 6).

For the marine phosphorus inventory, the precipitation of carbonate fluorapatite (CFA) in organic-rich sediments constitutes an important mode of removal of biologically available P from the ocean, accounting for some 10–15% of the total P removal⁹. Phosphogenesis in the contemporary ocean occurs mainly in the eastern tropical Pacific and Arabian Sea, and to a minor extent off Namibia and western Australia (Table 1). (The term phosphogenesis is used here exclusively with reference to relatively large P enrichments in margin sediments with porewater HPO₄²⁻ levels exceeding CFA saturation, although we recognize that

CFA occurs at trace levels elsewhere in deep-sea and margin sediments⁹. The marked spatial correspondence between phosphogenesis and water-column denitrification suggests that these processes are promoted by conditions prevailing in oxygen-deficient upwelling margins⁸.

The eastern tropical North Pacific hosts the most extensive coastal upwelling area and the largest mass of denitrifying waters in the world's oceans, and phosphogenic sediments are common in the bordering margins off Central America⁸. Off northwest Mexico, the oxygen minimum has near-zero oxygen concentrations extending from approximately 150 to 800 m water depth (Fig. 1a). Denitrification is a major respiration process in subsurface waters, as indicated by a large nitrate deficit in the water column that corresponds with the oxygen-deficient zone (Fig. 1b). Denitrification preferentially reduces ¹⁴N, and as a result the residual nitrate is progressively enriched in ¹⁵N in the denitrifying waters¹⁰ (Fig. 1c). This signal is transmitted to the sediments by upwelling of isotopically heavy nitrate into surface waters, its assimilation by phytoplankton, and the subsequent sinking of organic detritus to the sea floor^{6,11}. As a result, δ¹⁵N values in box cores are heavy, ranging between 9‰ and 10‰ across the margin⁸.

Contents of sedimentary organic carbon in box cores on the Mexican continental slope increase from 2 wt% on the outer shelf (shelf break ~150 m) to ~10% at 1,000 m depth (Fig. 1d). The sediments that are accumulating between 150 and ~800 m water depth are distinctly laminated, and contain high concentrations of P (Fig. 1e) owing to the presence of CFA as disseminated grains and discrete unconsolidated laminae typically < 1 cm thick. The CFA-enriched layers are also defined by large spikes in down-core P concentration profiles (plotted as maximum P contents in Fig. 1e), which cannot be accounted for by P contained in organic or lithogenous material. A close correspondence between increased interstitial dissolved HPO₄²⁻ concentrations and CFA precipitation can be seen in Fig. 1e and f. Thermodynamic calculations indicate that interstitial dissolved phosphate concentrations exceeding 40 μM should cause supersaturation with respect to CFA solubility in these sediments¹². The increased porewater P concentrations of phosphogenic sediments, which exceed CFA solubility by an order of magnitude, are attributed primarily to a large P supply to the sediments by way of settling biogenic debris and to some extent to the cycling of remineralized/redissolved P in suboxic sediments¹³. Although the roles of fish-bone dissolution, Fe-oxyhydroxide cycling and inhibited bioturbation in promoting CFA precipitation are being debated^{14,15}, phosphogenesis and porewater phosphate levels exceeding CFA saturation occur almost exclusively in organic-rich, suboxic upwelling margin sediments that underlie oxygen-deficient/denitrifying bottom waters (Fig. 1; refs 12–15).

A 10-m-long piston core (NH15P; 425 m water depth) raised from the upper slope of the Mexican margin provides a detailed history of productivity, denitrification and phosphogenesis on glacial–interglacial timescales (Fig. 2a–e). This core site is located where the oxygen minimum and denitrification are the most

Table 1 Estimated rates of burial of phosphogenic P in the modern ocean

Location	P burial*	Area (10 ⁶ km ²)
ETNP (off Mexico–Guatemala)†	0.3 to 1.6	0.5
ETSP (off Peru–Chile)†	0.8 to 6.3	0.2
Arabian Sea (off Oman and Indo-Pakistan) ¹⁵	0.076 to 1.04	0.4
Southwest Africa (Namibian shelf) ²³	2.8 to 3.3	0.03
West Australian shelf ²⁴	NA	<0.02
Global phosphogenic P burial‡	1.2 × 10 ¹⁰ mol yr ⁻¹	

* In units of μmol cm⁻² yr⁻¹, except where shown.

† Rates estimated from P contents in Mexican margin box cores (Fig. 1). Area calculated on the basis that phosphogenesis is occurring on the shelf and upper slope of the eastern North Pacific (ETNP), the eastern South Pacific (ETSP) and the Arabian Sea, and on the shelf off Namibia and West Australia.

‡ Estimated by extrapolating median P burial of 1 μmol cm⁻² yr⁻¹ over a total phosphogenic sediment area of 1.2 × 10⁶ km². NA, not available.

intense, and where CFA is precipitating in the modern sediments. X-radiographs reveal alternating cycles of laminated and bioturbated intervals down core that closely match glacial–interglacial cycles (Fig. 2a). Laminated sediments are confined to interglacial stages (1, 3 and 5), while glacial-age sediments are bioturbated (stages 2 and 4). Contents of organic carbon and opal are significantly higher in interglacial intervals compared to glacial sediments (Fig. 2c and d). Phosphorus-enriched layers are invariably confined to laminated interglacial sediments rich in organic carbon and opal (Fig. 2b–d). The lack of P-enrichment in glacial horizons suggests that phosphogenesis did not occur at the position of the modern oxygen minimum off northwest Mexico at these times.

Measurements of $\delta^{15}\text{N}$ in the core show cyclic variations, with particularly enriched values ($>8\text{‰}$) being largely confined to the interglacials (Fig. 2e). This suggests that large-scale denitrification occurred in this region only during the interglacials and was absent during the glacial periods^{6,11}. The close correspondence between the bioturbated (unlaminated) sections of the core and lighter $\delta^{15}\text{N}$ values indicates that waters overlying the core site were relatively well oxygenated and the sediments were less reducing during the deposition of these intervals. Such glacial–interglacial contrasts in the accumulation rates of organic carbon, opal, biogenic barium and $\delta^{15}\text{N}$ values are seen across the continental slope in this same area. This has led to the suggestion that biological productivity and upwelling rates diminished over the margin during the glacial periods^{6,16}. Relatively low export production during the glacial periods resulted in a decreased supply of settling biogenic debris and enhanced oxygenation of the water column and superficial sediments. These conditions in turn seem to have inhibited denitrification in the water column and sedimentary CFA precipitation.

These processes seem to have operated on other highly productive continental margins. Off Peru, sedimentary phosphorites in

laminated organic-rich sediments invariably formed during interglacial periods, and evidence for phosphogenesis is absent in glacial intervals^{17,18}—which are marked by reduced productivity and denitrification and increased oxygenation of sub-thermocline waters⁶. On the Arabian Sea margins, very high P contents are confined to interglacial sediments in contact with oxygen-deficient waters on the upper slope, and phosphogenesis seems to have decreased during glacial^{15,19} in response to reduced production and increased oxygenation of the water column^{6,20}.

The concomitant declines in phosphogenesis and denitrification during glacial periods should have had a much larger effect on the oceanic inventory of N than of P. Given that the continental margins bordering the eastern Pacific and the Arabian Sea account for more than 90% of the total area that supports phosphogenesis in the modern ocean (Table 1), the elimination of phosphogenesis in these regions during glacial stages should have resulted in a reduction in P burial of up to 10% of the modern total burial rate (about $8\text{--}12 \times 10^{10} \text{ mol P yr}^{-1}$; ref. 9). Assuming that the P removal rate in other oceanic regions is proportional to the P concentration in sea water, this would only account for an approximately 10% increase over the modern P inventory of the ocean ($3.2 \times 10^{15} \text{ mol P}$; ref. 9). In addition, the decline in P burial ($\sim 10\%$) due to diminished phosphogenesis is likely to have been compensated to some extent by increased P burial in glacial sediments in non-phosphogenic areas. Glacial–interglacial changes in reactive P input in the ocean—from the erosion of continental shelf sediments exposed during lowstands²¹, or from terrestrial biomass or soils—are poorly constrained. However, such changes are likely to be much smaller in magnitude, and the changes in P transfer rates at glacial–interglacial transitions much slower, than required to account for rapid changes in temperature and partial pressure of CO_2 recorded in ice cores⁶. Thus, the glacial increase in oceanic P inventory could not have been

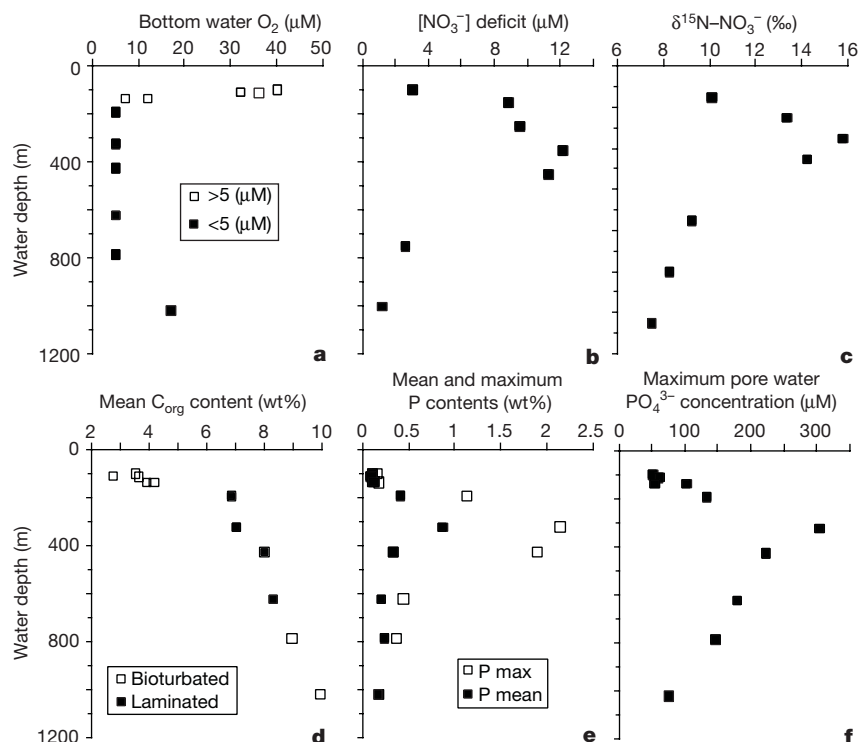


Figure 1 Water column and sediment properties across the continental shelf and slope off northwest Mexico. **a–c**, Water-column profiles of oxygen, nitrate deficit and the $\delta^{15}\text{N}$ of nitrate originally reported in ref. 8. **d**, Mean wt% of organic carbon (C_{org}) in box cores that were typically 15–30 cm long and were sectioned at 1–2-cm intervals. Down-core changes in organic carbon content are within 1 wt% in most box cores. **e**, Mean and maximum wt% P in the box cores plotted against water depth. The large difference

between mean and maximum P contents indicates the occurrence of CFA enriched laminae in sediments. **f**, Maximum phosphate concentrations in pore waters²⁵. Total carbon was determined by Carlo-Erba CNS analyser, and carbonate carbon by coulometry. Organic carbon was obtained by difference with a combined precision of $\pm 3.9\%$. P was determined using an automated Philips PW1400 X-ray fluorescence spectrometer with precision better than $\pm 5\%$.

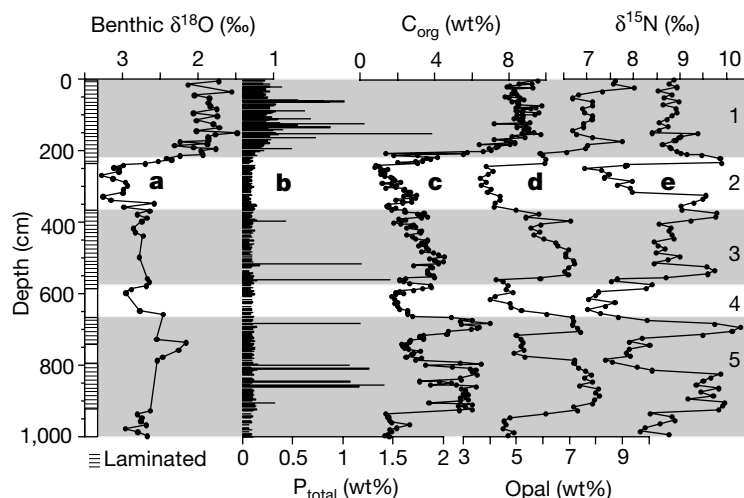


Figure 2 Records of $\delta^{18}\text{O}$ of benthic foraminifera, weight per cent phosphorus and organic carbon, and $\delta^{15}\text{N}$ of bulk sediments in Core NH15P (425 m water depth; $22^\circ 41.0' \text{ N}$, $106^\circ 28.8' \text{ W}$). Hatched bars on the left margin indicate the presence of visible laminations in X-radiographs. Isotope stage boundaries are based on 13 AMS- ^{14}C dates and benthic foraminiferal oxygen isotope stratigraphy¹⁶. Isotope stages are indicated near the right margin of the panels and interglacial stages are shaded grey. **a**, $^{18}\text{O}/^{16}\text{O}$ ratios were measured on *Bolivina* spp., using a VG PRISM mass spectrometer equipped with a VG Autocarb common-bath sample preparation system. The results are reported in the δ notation, $\delta^{18}\text{O} = [(^{18}\text{O}/^{16}\text{O})_{\text{sample}} / (^{18}\text{O}/^{16}\text{O})_{\text{standard}}] - 1$, in units of per mil, where the standard is PDB. The analytical precision for these measurements is $\pm 0.04\text{‰}$.

b, Phosphorus contents are presented as bar graphs with each bar representing a 2-cm sampling interval (see Fig. 1 legend for analytical details). Some of the most elevated P concentrations occur in intervals corresponding with layered structures, where the presence of CFA is identified by X-ray powder diffraction. Note that sample intervals with CFA laminae contain up to 3 wt% P, which corresponds to a CFA content of roughly 25% (pure CFA contains ~ 13 wt% P; ref. 26). The intervening sediments are also enriched in P

to varying degrees due to the presence of dispersed CFA grains. The P accumulation rate in the Holocene sediments of Core NH15P is on the order of $0.6 \mu\text{mol cm}^{-2} \text{ yr}^{-1}$ (calculated as the product of the sedimentation rate (17 cm kyr^{-1}), an average P content of 0.33 wt%, P atomic wt 30.97, and a dry bulk density of 0.35 g cm^{-3}). More than 80% of this is in the form of reactive P (CFA + organic P), based on comparison of the P/Al ratio in CFA-free and organic-free detritus (~ 0.0088 ; ref. 26) and the mean in the Holocene section of NH15P (0.044). In contrast, P burial in the Last Glacial Maximum (LGM) sediments in Core NH15P was approximately $0.1 \mu\text{mol cm}^{-2} \text{ yr}^{-1}$ (4 cm kyr^{-1} ; 0.09 wt% P; 0.7 g cm^{-3} ; 0.013 P/Al), of which the dominant host, based on normative calculations, is lithogenic detritus. **c**, Organic carbon determination is described in Fig. 1 legend. **d**, Opal determinations were made by colorimetry after leaching sediments with 2 M Na_2CO_3 at 85°C for 3 h. Analytical precision is $\pm 6\%$. **e**, Nitrogen isotope ratios were determined using a Carlo-Erba CHN analyser interfaced directly to the mass spectrometer. $\delta^{15}\text{N} = [(^{15}\text{N}/^{14}\text{N})_{\text{sample}} / (^{15}\text{N}/^{14}\text{N})_{\text{standard}}] - 1$, in units of per mil, where the standard is atmospheric nitrogen; the precision of this measurement is better than $\pm 0.3\text{‰}$.

more than 10% of the present value, whereas the glacial increase in fixed nitrogen due to reduced water-column denitrification was probably substantial (about 50% of the modern value; see, for example, ref. 6). In addition, the oceanic residence time for N (3 kyr; ref. 4) is much shorter than that for P (20–30 kyr; ref. 9), and therefore the N content could rise much faster than that of P during glacial periods.

On the basis of the above inferences, we propose two models for nutrient contents of the glacial ocean. First, the rising glacial dissolved N/P ratio rapidly shifted ecological advantage from nitrogen-fixers (diazotrophs) to non-nitrogen-fixing algae¹, counteracting the effect of decreased denitrification and effectively anchoring the N to the P inventory. This would restore the Redfield ratio. Implicit in this argument, however, is the assumption that the ecological response by nitrogen fixation is as immediate as the increase in N/P ratio. The second model involves slow compensation by nitrogen fixation for variations in N/P ratios, as has been suggested recently⁵. In this case, glacial N contents would have rapidly increased until the N/P ratio was high enough to subjugate nitrogen fixation, allowing the seawater N/P ratio to gradually return to its Redfield value.

But regardless of which of the above two models is correct, they allow two important conclusions to be drawn. First, nitrogen fixation in glacial times is likely to have been lower (and not higher) than interglacial nitrogen fixation—owing to P limitation (higher N/P)—and therefore the ice-age flux of aerosol iron to oligotrophic regions of the sea is unlikely to have been as biogeochemically important as suggested previously^{1,2}. This is consistent with recent observations that nitrogen fixation could be phosphorus-limited in Fe-replete conditions²². Second, glacial increases in ocean productivity resulting from changes in nutrient inventory

and their influences on the partial pressure of CO_2 are likely to have been small, having been fundamentally dictated by the slight and slow increase in P inventory. □

Received 25 August; accepted 16 November 2001.

- Falkowski, P. G. Evolution of nitrogen cycle and its influence on the biological pump in the ocean. *Nature* **342**, 637–642 (1997).
- Broecker, W. S. & Henderson, G. M. The sequence of events surrounding Termination II and their implications for the cause of glacial-interglacial CO_2 changes. *Paleoceanography* **13**, 352–364 (1998).
- Redfield, A. C., Ketchum, B. H. & Richards, F. A. in *The Sea* Vol. 2 (ed. Hill, M. N.) 36–77 (McGraw Hill, London, 1963).
- Tyrrell, T. The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature* **400**, 525–531 (1999).
- Hood, R. R. et al. Answers sought to the enigma of marine nitrogen fixation. *Eos* **81**, 1–3 (2000).
- Ganeshram, R. S., Pedersen, T. F., Calvert, S. E., McNeil, G. W. & Fontugne, M. R. Glacial-interglacial variability in denitrification in the world's oceans: Causes and consequences. *Paleoceanography* **15**, 361–376 (2000).
- Codispoti, L. A. & Christensen, J. P. Nitrification, denitrification and nitrous oxide cycling in the eastern tropical South Pacific Ocean. *Mar. Chem.* **16**, 277–300 (1985).
- Codispoti, L. A. in *Productivity in the Ocean: Past and Present* (eds Berger, W. H., Smetacek, V. S. & Wefer, G.) 377–394 (Wiley & Sons, Chichester, UK, 1989).
- Delaney, M. L. Phosphorus accumulation in the marine sediments and the oceanic phosphorus cycle. *Glob. Biogeochem. Cycles* **12**, 563–572 (1998).
- Brandes, J. A., Devol, A. H., Yoshinari, T., Jayakumar, D. A. & Naqvi, S. W. A. Isotopic composition of nitrate in the central Arabian Sea and eastern tropical North Pacific: a tracer for mixing and nitrogen cycles. *Limnol. Oceanogr.* **43**, 1680–1689 (1998).
- Ganeshram, R. S., Pedersen, T. F., Calvert, S. E. & Murray, J. W. Large changes in oceanic nutrient inventories from glacial to interglacial periods. *Nature* **376**, 755–758 (1995).
- Jahnke, R. A., Emerson, S. R., Roe, K. V. & Burnett, W. C. Present day formation of apatite on the Mexican continental margin. *Geochim. Cosmochim. Acta* **47**, 259–266 (1983).
- Froelich, P. N. et al. Early diagenesis of organic matter in Peru continental margin sediments: Phosphorite precipitation. *Mar. Geol.* **80**, 309–343 (1988).
- Schuffert, J. D., Kastner, M. & Jahnke, R. A. Carbon and phosphorus burial associated with modern phosphorite formation. *Mar. Geol.* **146**, 21–23 (1998).
- Schenau, S. J., Slomp, C. P. & De Lange, G. Phosphogenesis and active phosphorite formation in sediments from the Arabian Sea oxygen minimum zone. *Mar. Geol.* **169**, 1–19 (2000).
- Ganeshram, R. S. & Pedersen, T. F. Glacial-interglacial variability in upwelling and bioproductivity off N. W. Mexico. *Paleoceanography* **13**, 634–645 (1998).

17. Burnett, W. C. Apatite-glaucinite associations off Peru and Chile: palaeo-oceanographic implications. *J. Geol. Soc. Lond.* **137**, 757–764 (1980).
18. Garrison, R. E. & Kastner, M. Phosphatic sediments and rocks recovered from Peru margin during ODP Leg 112. *Proc. ODP Sci. Res.* **112**, 111–134 (1990).
19. Shimmield, G. B. & Mowbray, S. R. The inorganic geochemical record of the Northwest Arabian Sea. *Proc. ODP Sci. Res.* **117**, 409–429 (1991).
20. Altabet, M. A., François, R., Murray, D. W. & Prell, W. L. Climate-related variations in denitrification in the Arabian Sea from $^{15}\text{N}/^{14}\text{N}$ ratios. *Nature* **373**, 506–509 (1995).
21. Ruttenger, K. C. Reassessment of the oceanic residence time of phosphorus. *Chem. Geol.* **107**, 405–409 (1993).
22. Sañudo-Wilhelmy, S. A. *et al.* Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic Ocean. *Nature* **411**, 66–69 (2001).
23. Filippelli, G. M. Controls on phosphorus concentrations and accumulation in oceanic sediments. *Mar. Geol.* **139**, 231–240 (1997).
24. O'Brien, G. W. & Heggge, D. East Australian continental margin phosphorites. *Eos* **69**, 2 (1988).
25. Lambourn, L. D., Devol, A. H. & Murray, J. W. *R/V New Horizon 90-5 Cruise: Water Column and Porewater Data* (Special report no. 110, School of Oceanography, University of Washington, 1991).
26. Price, N. B. & Calvert, S. E. The geochemistry of phosphorites from the Namibian Shelf. *Chem. Geol.* **23**, 151–170 (1978).

Acknowledgements

We thank A. Devol and J. Murray for the opportunity to collect sediment cores during the NSF-supported voyage 90-5 of RV *New Horizon* to the Mexican margin. We thank M. Soon and B. Nielsen for carrying out elemental and isotopic determinations, and J. Brandes for permission to use his water column nitrogen isotope data. This work was supported by the Natural Sciences and Engineering Research Council of Canada and the Natural Environment Research Council of the UK.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.S.G. (e-mail: Raja.Ganeshram@glg.ed.ac.uk).

The effect of millennial-scale changes in Arabian Sea denitrification on atmospheric CO₂

Mark A. Altabet*, Matthew J. Higginson* & David W. Murray†

* School for Marine Science and Technology, University of Massachusetts Dartmouth, 706 S. Rodney French Boulevard, New Bedford, Massachusetts 02744-1221, USA

† Center for Environmental Studies, Box 1943, Brown University, Providence, Rhode Island 02912, USA

Most global biogeochemical processes are known to respond to climate change, some of which have the capacity to produce feedbacks through the regulation of atmospheric greenhouse gases¹. Marine denitrification—the reduction of nitrate to gaseous nitrogen—is an important process in this regard, affecting greenhouse gas concentrations directly through the incidental production of nitrous oxide, and indirectly through modification of the marine nitrogen inventory and hence the biological pump for CO₂. Although denitrification has been shown to vary with glacial–interglacial cycles^{2,3}, its response to more rapid climate change has not yet been well characterized. Here we present nitrogen isotope ratio, nitrogen content and chlorin abundance data from sediment cores with high accumulation rates on the Oman continental margin that reveal substantial millennial-scale variability in Arabian Sea denitrification and productivity during the last glacial period. The detailed correspondence of these changes with Dansgaard–Oeschger events recorded in Greenland ice cores⁴ indicates rapid, century-scale reorganization of the Arabian Sea ecosystem in response to climate excursions, mediated through the intensity of summer monsoonal upwelling. Considering the several-thousand-year residence time of fixed

nitrogen in the ocean, the response of global marine productivity to changes in denitrification would have occurred at lower frequency and appears to be related to climatic and atmospheric CO₂ oscillations observed in Antarctic ice cores between 20 and 60 kyr ago⁵.

Denitrification occurs under suboxic conditions when bacteria utilize NO₃[−] as an electron acceptor and in doing so convert it primarily to N₂ gas. This process is the primary loss mechanism for combined nitrogen from the biosphere, and thus has an important role for the nitrogen cycle and the biogeochemical cycles linked to it. In the ocean, denitrification occurs in organic-rich continental margin sediments and in intermediate waters within intense oxygen–minimum zones. Of the latter, the Arabian Sea supports approximately one-third of marine water-column denitrification⁶. Denitrification also strongly fractionates nitrogen isotopes, leaving the remaining NO₃[−] enriched in ¹⁵N (refs 7 and 8). A palaeoceanographic record for denitrification intensity is created when ¹⁵N-enriched NO₃[−] is transported to surface waters and consumed by phytoplankton, with subsequent downward transport of organic matter and preservation in the sediments. When there is good-to-excellent preservation of organic matter at the sea floor due to either suboxic bottom water and/or high sediment accumulation rates, sediment δ¹⁵N faithfully records the δ¹⁵N of sinking organic matter⁹.

Previous studies have shown strong coupling between variations in denitrification intensity and climate change at orbital periodicities (> 19 kyr) for the Arabian Sea^{2,3}, as well as during glacial–interglacial transitions in other major regions for water-column denitrification¹⁰. Here we take advantage of two deep-sea cores from the Oman margin (Fig. 1) with relatively high sediment accumulation rates (~15 cm kyr^{−1}) to examine possible high-frequency variability in denitrification and productivity and their relationship to climate change. Sampling every 1 cm achieved an

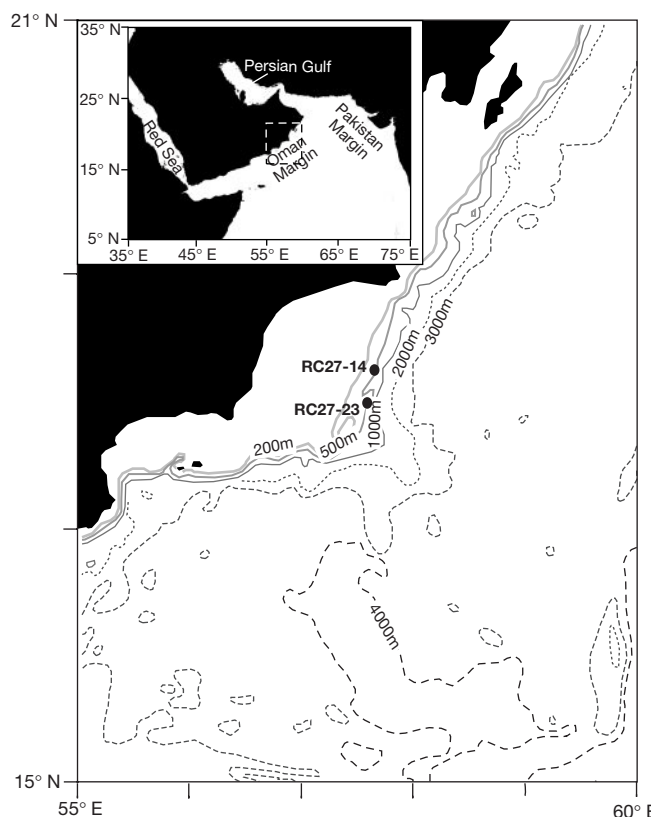


Figure 1 Location of sediment cores on the Oman margin, and (inset) the location of the Oman margin within the Arabian Sea. Both sites (RC27-14: 18° 15.2' N, 57° 39.3' E, 596 m water depth; RC27-23: 17° 59.6' N, 57° 35.4' E, 820 m water depth) lie at depths inside the oxygen–minimum zone, which impinges on the continental slope between 200 and 1,200 m water depth.

average temporal resolution of <100 yr over the past 65 kyr. In addition, these cores are located at depths within the modern oxygen-minimum zone, ensuring sufficient preservation of buried organic matter and preventing post-depositional modification of $\delta^{15}\text{N}$ values.

High-resolution $\delta^{15}\text{N}$ analysis reveals large century- to millennial-scale oscillations over the past 65 kyr in both RC27-14 and RC27-23 (Fig. 2), and these two records are in detail nearly identical. A coarse timescale for both cores is based on foraminiferal $\delta^{18}\text{O}$, which records major climate stage transitions in the vicinity of 20 kyr ago (marine isotope stage 2) and 60 kyr ago (marine isotope stage 4)¹¹. The higher-frequency oscillation in $\delta^{15}\text{N}$, however, is remarkably similar to the Greenland ice-core record of changes in polar climate (GISP2 $\delta^{18}\text{O}$)^{12,13}. This similarity extends well beyond the number of events, and encompasses their relative duration and shape. Uniquely for a marine record, a saw-tooth pattern of rapid transition to elevated $\delta^{15}\text{N}$ followed by a more gradual decrease is also apparent for the longer-duration interstadials (for example, Dansgaard–Oeschger events 12 and 14; Fig. 2). Accordingly, a high-resolution time scale was developed by assuming that Oman margin $\delta^{15}\text{N}$ and GISP2 $\delta^{18}\text{O}$ were in phase and synchronous. A series of 45

correlation points (as indicated in Fig. 2) were identified for each core, and GISP2 chronology was transferred to the Oman margin cores by linear interpolation between them¹⁴. Not surprisingly, spectral analysis of the resulting $\delta^{15}\text{N}$ time series reveals the dominant 1,460-year periodicity (first recognized from the Greenland ice cores and marine records of ice-rafted debris from the northern North Atlantic¹⁵) as well as other periodicities (1,020, 2,300, 3,400, 6,700 yr). Both cores also show very similar and near-linear depth versus age relationships (see Supplementary Information), indicating that our assumption of synchronicity with Greenland ice-core $\delta^{18}\text{O}$ is not unrealistic. Finally, our ice-core-based age scale is not violated by foraminiferal ^{14}C dates¹¹ (within error margins) converted to calendar years on the basis of the latest available calibration and assuming an unchanging marine reservoir ^{14}C age of 400 yr (ref. 16). Uncertainties associated with the assumed reservoir age, calendar age calibration, and the GISP2 timescale itself prevent further chronological refinement at present. Whether there are phase shifts between the GISP2 $\delta^{18}\text{O}$ and Oman margin $\delta^{15}\text{N}$ records remains an open question, awaiting advances in the dating of marine sediments.

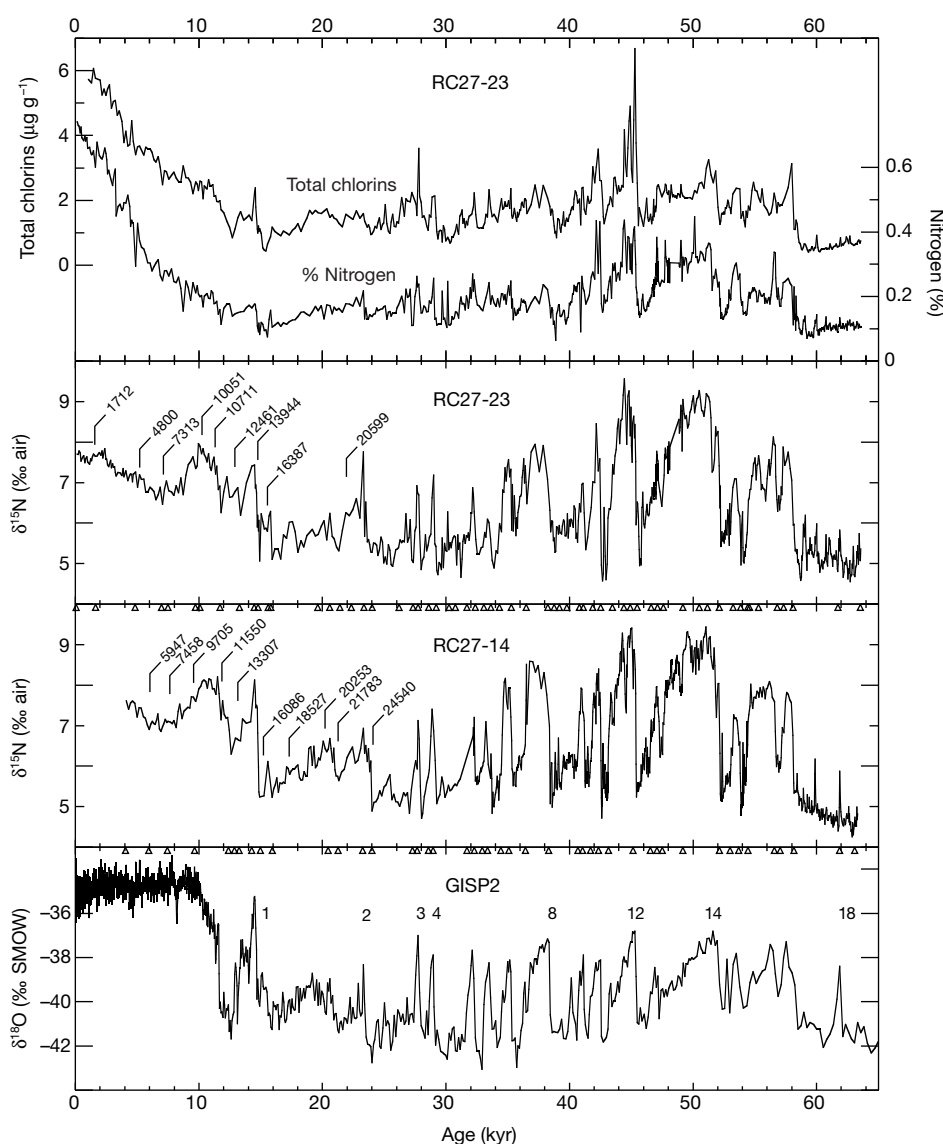


Figure 2 Time series of sediment-core data along with the GISP2 $\delta^{18}\text{O}$ record for the past 65,000 years. The figure shows the correlation of high-resolution $\delta^{15}\text{N}$ records for denitrification from cores RC27-14 and RC27-23, and nitrogen content and chlorin palaeoproductivity proxy data for core RC27-23, with the GISP2 $\delta^{18}\text{O}$ record for the past 65,000 years. Chronologies were constructed by correlating mid-points of stadal–

interstadial transitions (indicated with open triangles) with those in the GISP2 record¹⁴. AMS (accelerator mass spectrometry) ^{14}C dates¹¹, calibrated to a calendar timescale, are shown for comparison. SMOW, standard mean ocean water. Numbers on the bottom panel refer to Dansgaard–Oeschger events.

The strong correspondence in frequency and pattern between the GISP2 $\delta^{18}\text{O}$ and Oman margin $\delta^{15}\text{N}$ records is evidence for a millennial-scale mechanistic linkage between Northern Hemisphere climate and Arabian Sea denitrification. $\delta^{15}\text{N}$ and thus denitrification is near modern, high values during the warm phases of the Dansgaard–Oeschger events (interstadials), and at near-minimal glacial values during the cold phases of these events (stadials). Many excursions in denitrification were at times as abrupt as those for high-latitude Northern Hemisphere climate, with sub-century responses as rapid as can be resolved by our sampling resolution. Denitrification requires suboxic conditions, and thus depends on the intensity and extent of the subsurface oxygen–minimum zone. This will, in turn, vary with (1) local productivity and downward flux of organic carbon and (2) the ventilation age of Arabian Sea intermediate–water sources^{17,18}. Previous evidence for a link between climate excursions and variations in Arabian Sea productivity and suboxia has come from down-core organic carbon abundances, faunal abundances and lamination patterns on the Pakistan margin^{19,20}, and very recently lower-resolution $\delta^{15}\text{N}$ (ref. 21), indicating the regional nature of these events. For RC27-14 and RC27-23, sediment nitrogen content and chlorin concentration are

here used as proxies for palaeoproductivity²². Both records indicate that elevated denitrification during Dansgaard–Oeschger events was associated with increased productivity. Sediment nitrogen content and chlorin concentration do not follow the GISP2 record as closely as does $\delta^{15}\text{N}$, but this difference may reflect the extrinsic nature of these proxies. Strictly speaking, the accumulation rate (rather than the concentration) of nitrogen or chlorins should most closely follow productivity. However, rapid variations in sediment accumulation rate at millennial periodicities are difficult to resolve, and such uncertainties dominate variations in the accumulation rate of these proxies. Nevertheless, productivity changes clearly had a primary role in forcing variations in Arabian Sea denitrification at the timescale of Dansgaard–Oeschger events.

High primary production and downward particle flux are largely the result of strong summertime monsoon winds, which force upwelling of nutrient-rich water in the modern Arabian Sea²³. Wind strength proxies (grain size) in the sediment record have been previously shown to vary at orbital timescales, and can account for climate-linked variations in productivity²⁴. This mechanistic coupling of climate to Arabian Sea denitrification probably also continues at the millennial scales observed here. Specifically,

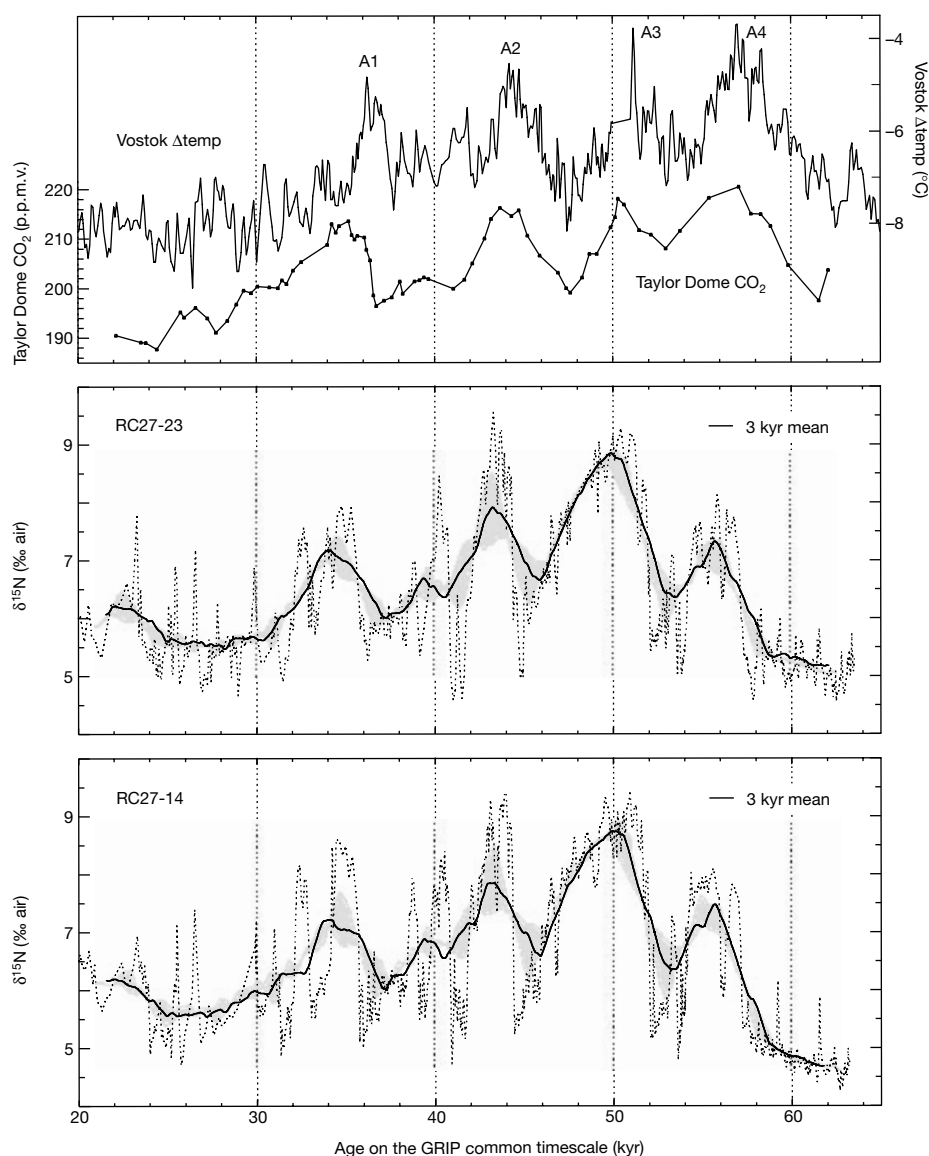


Figure 3 The 3-kyr moving average of the $\delta^{15}\text{N}$ record compared to Antarctic ice-core records of climate and atmospheric CO_2 (ref. 5). The ± 1 kyr envelope is shown shaded. Variations in Arabian Sea denitrification would alter the marine combined N inventory on

timescales similar to its 3-kyr residence time, in the absence of other effects. Δtemp , the relative temperature change at Vostok derived from ice D/H ratios, A1–A4, Antarctic climate events.

meteorological forcing of productivity and denitrification can explain the apparent rapid increase of Arabian Sea denitrification associated with the onset of each Dansgaard–Oeschger warm phase as observed in Greenland ice cores. Changing high-latitude temperature occurred at the same time as altered weather patterns linked to the strength of the summer south Asian monsoon. For example, Dansgaard–Oeschger events coincided with latitudinal shifts in the intertropical convergence zone (ITCZ), which may have been their source²⁵. Through atmospheric forcing, Northern Hemisphere climate events would have a practically instantaneous expression in the Arabian Sea denitrification record.

The close linkage between Dansgaard–Oeschger events and Arabian Sea denitrification has implications for our understanding of Earth system dynamics. The $\delta^{15}\text{N}$ record demonstrates the high sensitivity of the Arabian Sea oxygen-minimum zone to Northern Hemisphere climate change on short timescales. Our records show clear instances (Fig. 2) where Arabian Sea intermediate waters were transformed, in a century or less, from an oxic state to a suboxic state that was even more intense than at present. These events must have produced marked alterations of the Arabian Sea ecosystem with respect to benthic and mesopelagic organisms dependent on O_2 . As these changes in subsurface oxygenation appear to be forced by oscillation in upwelling-stimulated productivity, large shifts are also implied in populations at trophic levels, ranging from zooplankton to marine mammals. These observations suggest potential for similarly rapid alteration of marine ecosystems in response to future climate change. In this regard, the western North American margin may be particularly vulnerable, as this region is dominated by coastal upwelling and has subsurface water that is at present low in O_2 —but not so low as to be suboxic.

A number of important biogeochemical processes are restricted to suboxic zones. For example, fluxes of N_2O and CH_4 to the atmosphere are relatively enhanced in the modern Arabian Sea²⁶. During cold phases of the Dansgaard–Oeschger events with apparent elimination of suboxic conditions, these fluxes would have been expected to be substantially reduced²¹. Similarly, the modern Arabian Sea is a major site for water-column denitrification. Recently estimated at 33 Tg yr^{-1} (ref. 6), Arabian Sea water-column denitrification accounts for approximately 15% of the oceanic sink for combined nitrogen²⁷. In isolation, millennial-scale modulation of Arabian Sea denitrification would have produced corresponding oscillations in the budget for marine combined N.

The frequency spectrum for any variation in the inventory of marine combined N would be constrained by its residence time, which at present is about 3 kyr. Rectifying the $\delta^{15}\text{N}$ record with a 3-kyr moving average produces a time series that inversely describes the corresponding changes in marine nitrogen inventory due to oscillations in Arabian Sea denitrification (Fig. 3). Changes in marine N inventory would alter the residence time, but consideration of a ± 1 kyr envelope (+25% to –33% in inventory) does not show a significant difference in result. The 3-kyr moving average for $\delta^{15}\text{N}$ exhibits a series of maxima and minima between 25 and 65 kyr ago which are strikingly similar to those found in Antarctic ice-core temperature and atmospheric CO_2 records⁵. At present, it is difficult to assess the phase relationship between these records, but they are strongly suggestive of a mechanistic linkage.

Change in oceanic nutrient inventory is one means for forcing changes in atmospheric CO_2 via alteration of marine productivity^{28,29}. Increased denitrification would reduce combined N and thus the strength of the marine biological pump for carbon if sustained for a sufficiently long period of time. This model requires substantial increases in the oceanic ratio of N to P above the canonical Redfield ratio such as found in the modern Mediterranean Sea³⁰. Variations in Arabian Sea denitrification between 25 and 65 kyr ago may thus have contributed to forcing the moderate variations in atmospheric CO_2 at that time. If so, the marine N cycle with a several-thousand-year residence time may provide the filter

translating the higher-frequency Dansgaard–Oeschger events into the lower-frequency CO_2 signal and global climate change recorded in the Southern Hemisphere. □

Received 18 April; accepted 27 November 2001.

- Kennett, J. P., Cannariato, K. G., Hendy, I. L. & Behl, R. J. Carbon isotopic evidence for methane hydrate instability during Quaternary interstadials. *Science* **288**, 128–133 (2000).
- Altabet, M. A., Francois, R., Murray, D. W. & Prell, W. L. Climate-related variations in denitrification in the Arabian Sea from sediment $^{15}\text{N}/^{14}\text{N}$ ratios. *Nature* **373**, 506–509 (1995).
- Altabet, M. A., Murray, D. W. & Prell, W. L. Climatically linked oscillations in Arabian Sea denitrification over the past 1 m.y.: Implications for the marine N cycle. *Paleoceanography* **14**, 732–743 (1999).
- Dansgaard, W. *et al.* A new Greenland deep ice core. *Science* **218**, 1273–1277 (1982).
- Indermühle, A., Monnin, E., Stauffer, B., Stocker, T. F. & Wahlen, M. Atmospheric CO_2 concentration from 60 to 20 Kyr BP from the Taylor Dome ice core, Antarctica. *Geophys. Res. Lett.* **27**, 735–738 (2000).
- Howell, E. A., Doney, S. C., Fine, R. A. & Olson, D. B. Geochemical estimates of denitrification in the Arabian Sea and the Bay of Bengal during WOCE. *Geophys. Res. Lett.* **24**, 2549–2552 (1997).
- Cline, J. D. & Kaplan, I. R. Isotopic fractionation of dissolved nitrate during denitrification in the Eastern Tropical North Pacific Ocean. *Mar. Chem.* **3**, 271–299 (1975).
- Brandes, J. A., Devol, A. H., Yoshinari, T., Jayakumar, D. A. & Naqvi, S. W. A. Isotopic composition of nitrate in the central Arabian sea and eastern tropical North Pacific: a tracer for mixing and nitrogen cycles. *Limnol. Oceanogr.* **43**, 1680–1689 (1998).
- Altabet, M. A. *et al.* The nitrogen isotope biogeochemistry of sinking particles from the margin of the Eastern North Pacific. *Deep-Sea Res.* **46**, 655–679 (1999).
- Ganeshram, R. S., Pedersen, T. F., Calvert, S. E., McNeill, G. W. & Fontugne, M. R. Glacial-interglacial variability in denitrification in the world's oceans: Causes and consequences. *Paleoceanography* **15**, 361–376 (2000).
- Anderson, D. M. *Foraminifer Evidence of Monsoon Upwelling off Oman during the Later Quaternary*. Thesis, Brown Univ. (1991).
- Stuiver, M. & Grootes, P. M. GISP2 oxygen isotope ratios. *Quat. Res.* **53**, 277–284 (2000).
- Grootes, P. M. & Stuiver, M. Oxygen 18/16 variability in Greenland snow and ice with 10^3 to 10^5 -year time resolution. *J. Geophys. Res.* **102**, 26455–26470 (1997).
- Shackleton, N. J., Hall, M. A. & Vincent, E. Phase relationships between millennial-scale events 64,000–24,000 years ago. *Paleoceanography* **15**, 565–569 (2000).
- Bond, G. *et al.* A pervasive millennial-scale cycle in North Atlantic Holocene and glacial climates. *Science* **278**, 1257–1266 (1997).
- Stuiver, M., Reimer, P. J. & Braziunas, T. F. High-precision radiocarbon age calibration for terrestrial and marine samples. *Radiocarbon* **40**, 1127–1151 (1998).
- Warren, B. A. in *Biogeochemistry of the Arabian Sea: Present Information and Gaps* (ed. Lal, D.) 301–314 (Indian Academy of Sciences, Bangalore, India, 1994).
- Olson, D. B., Hitchcock, G. L., Fine, R. A. & Warren, B. A. Maintenance of the low-oxygen layer in the central Arabian Sea. *Deep-Sea Res.* **40**, 673–685 (1993).
- Schulz, H., von Rad, U. & Erlenkeuser, H. Correlation between Arabian Sea and Greenland climate oscillations of the past 110,000 years. *Nature* **393**, 54–57 (1998).
- von Rad, U. *et al.* Multiple monsoon-controlled breakdown of oxygen-minimum conditions during the past 30,000 years documented in laminated sediments off Pakistan. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **152**, 129–161 (1999).
- Suthhof, A., Ittekkot, V. & Gaye-Haake, B. Millennial-scale oscillation of denitrification intensity in the Arabian Sea during the late Quaternary and its potential influence on atmospheric N_2O and global climate. *Glob. Biogeochem. Cycles* **15**, 637–650 (2001).
- Higginson, M. J. *Chlorin Pigment Stratigraphy as a New and Rapid Palaeoceanographic Proxy in the Quaternary*. Thesis, Univ. Bristol (2000).
- Honjo, S., Dymond, J., Prell, W. & Ittekkot, V. Monsoon-controlled export fluxes to the interior of the Arabian Sea: U.S. JGOFS 1994–1995 experiment. *Deep-Sea Res.* **46**, 1859–1902 (1999).
- Clemens, S., Prell, W., Murray, D., Shimmield, G. & Weedon, G. Forcing mechanisms of the Indian Ocean monsoon. *Nature* **353**, 720–725 (1991).
- Peterson, L. C., Haug, G. H., Hughes, K. A. & Röhl, U. Rapid changes in the hydrologic cycle of the tropical Atlantic during the last glacial. *Science* **290**, 1947–1951 (2000).
- Upstill-Goddard, R. C., Barnes, J. & Owens, N. J. P. Nitrous oxide and methane during the 1994 SW monsoon in the Arabian Sea/northwestern Indian Ocean. *J. Geophys. Res.* **104**, 30067–30084 (1999).
- Gruber, N. & Sarmento, J. L. Global patterns of marine nitrogen fixation and denitrification. *Glob. Biogeochem. Cycles* **11**, 235–266 (1997).
- Broecker, W. S. Glacial to interglacial changes in ocean chemistry. *Prog. Oceanogr.* **11**, 151–197 (1982).
- Sigman, D. M. & Boyle, E. A. Glacial/interglacial variations in atmospheric carbon dioxide. *Nature* **407**, 859–869 (2000).
- Krom, M. D., Brenner, S., Kress, N., Neori, A. & Gordon, L. I. Nutrient distributions during an annual cycle across a warm-core eddy from the E Mediterranean Sea. *Deep-Sea Res.* **40**, 805–825 (1993).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank R. Singh and X. Huang for technical assistance. Sediment core material was obtained from the collection and curating facilities of the Lamont–Doherty Earth Observatory core collection with support from the US NSF and ONR. We thank S. Clemens, L. Labeyrie and J. Kennett for discussions. This work was supported by the US NSF ESH program and the Camille and Henry Dreyfus Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.A.A. (e-mail: maltabet@umassd.edu).

Parasitic Cape honeybee workers, *Apis mellifera capensis*, evade policing

Stephen J. Martin*, Madeleine Beekman*†, Theresa C. Wossler‡ & Francis L. W. Ratnieks*

* Laboratory of Apiculture and Social Insects, Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

‡ Department of Zoology and Entomology, University of Pretoria, Pretoria 0002, South Africa

† Present address: Schools of Biological Sciences & Mathematics and Statistics, University of Sydney, Sydney NSW 2006, Australia

Relocation of the Cape honeybee, *Apis mellifera capensis*, by bee-keepers from southern to northern South Africa in 1990 has caused widespread death of managed African honeybee, *A. m. scutellata*, colonies¹. *Apis mellifera capensis* worker bees are able to lay diploid, female eggs without mating by means of automictic thelytoky² (meiosis followed by fusion of two meiotic products to restore egg diploidy), whereas workers of other honeybee subspecies are able to lay only haploid, male eggs. The *A. m. capensis* workers, which are parasitizing and killing *A. m. scutellata* colonies in northern South Africa, are the asexual offspring of a single, original worker in which the small amount of genetic variation observed is due to crossing over during meiosis³ (P. Kryger, personal communication). Here we elucidate two principal mechanisms underlying this parasitism. Parasitic *A. m. capensis* workers activate their ovaries in host colonies that have a queen present (queenright colonies), and they lay eggs that evade being killed by other workers (worker policing)—the normal fate of worker-laid eggs in colonies with a queen^{4–8}. This unique parasitism by workers is an instance in which a society is unable to control the selfish actions of its members.

In a honeybee (*Apis mellifera*) colony, reproduction is monopolized by the single queen. In queenright colonies of European subspecies only a few workers⁶ (approximately 0.01% of about 30,000 in a colony) have fully active ovaries, and most of their eggs are policed by other workers^{5–8}. (Worker policing is selectively favoured because workers are more related to the sons of their mother/queen (0.25) than to the sons of their sister workers (~0.15)^{4,7}.) We compared ovary activation and the acceptance of eggs laid by *A. m. capensis* workers parasitizing native *A. m. scutellata* colonies in the Pretoria region of northern South Africa, with *A. m. scutellata* workers and queens. Newly emerged *A. m. capensis* workers introduced into a queenright *A. m. scutellata*

Table 1 Numbers of worker-laid eggs in *A. m. scutellata* colonies

Day	Colony				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	2	0	0
3	0	0	0	0	1
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	1	0	1	0
8	0	0	0	0	0
Total	0	1	2	1	1

Eggs were laid in drone cells to which the queen was denied access. All of the eggs laid ($n = 5$) had been removed by workers (worker policing) before inspection on the next day. (See Methods for details of the experimental design.)

colony all activated their ovaries, whereas none of the *A. m. scutellata* workers did so (Fig. 1). Furthermore, eggs laid by *A. m. capensis* workers are not effectively removed by worker policing (Fig. 2). These eggs are nearly as acceptable as eggs laid by *A. m. scutellata* queens, possibly through chemical mimicry^{6,9}. In contrast, eggs laid by *A. m. scutellata* workers are policed effectively with almost none remaining after 20 h. This demonstrates that the parasitic *A. m. capensis* workers actually evade worker policing rather than exploit a situation where worker policing is lacking. An alternative hypothesis for the lack of policing, that all *A. m. capensis* worker-laid eggs evade policing, is also rejected because eggs laid by most *A. m. capensis* workers are policed in *A. m. scutellata* discriminator colonies (P. Neumann, C. W. W. Pirk and F.L.W.R., unpublished data). A second alternative hypothesis is that workers simply discriminate between diploid (queen-laid and *A. m. capensis* worker-laid) and haploid (normal worker-laid) eggs; however, studies of European honeybee subspecies have shown that this does not occur¹⁰.

Occasional worker-laid eggs were found in five *A. m. scutellata* colonies in cells above the queen excluder—a part of the hive that the queen cannot enter. All of these eggs were gone 24 h later (Table 1). This is very similar to what occurs in European subspecies of *Apis mellifera*^{6,8}, and further confirms that both egg laying by workers and worker policing occur at low rates in normal *A. m. scutellata* colonies.

Our results show that the parasitic *A. m. capensis* workers have at least two traits enabling them to reproduce in queenright *A. m. scutellata* colonies. The first is the ability to activate their ovaries in a queenright colony, and the second is the ability to lay eggs that evade worker policing. In this respect they resemble the anarchistic worker honeybees that occasionally occur in European subspecies^{7,10,11}; however, anarchistic workers lay haploid eggs that develop into males. By laying thelytokous diploid eggs the *A. m. capensis* workers do not merely reproduce, they replicate;

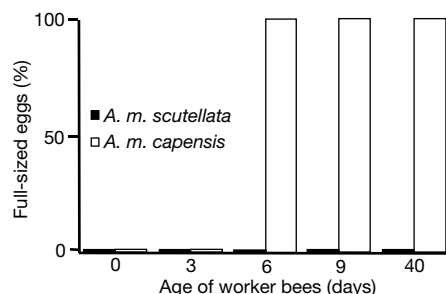


Figure 1 Ovary activation in *A. m. scutellata* and parasitic *A. m. capensis* workers fostered in a queenright *A. m. scutellata* colony. One hundred each of newly emerged *A. m. scutellata* and parasitic *A. m. capensis* workers were marked and introduced into a

queenright *A. m. scutellata* colony. Ten workers of each type were dissected on introduction and after 3, 6, 9 and 40 days to determine if full-sized eggs were present in their ovaries.

that is, they produce more parasitic workers. As a result the parasitic *A. m. capensis* workers increase in number within a host colony. This leads to the death of the host colony on which they depend. An important factor causing the death of a colony seems to be the dwindling numbers of *A. m. scutellata* workers that perform foraging duties (*A. m. capensis* workers are greatly under-represented in the foraging force of an infected colony) owing to death of the queen, and, before queen death, competition for egg laying between *A. m. capensis* workers and the queen¹³. Unfortunately, bee-keeping practices such as collecting wild swarms, seasonal movement of colonies by truck, and the close siting of hives in apiaries facilitate the horizontal spread of the *A. m. capensis* workers between colonies and provide a continuing source of hosts.

The mechanisms enabling intracolony parasitism that we document above are necessary for the parasitism to occur because worker policing is normally extremely efficient at killing worker-laid eggs, and because normal workers rarely activate their ovaries in queenright colonies^{5–8}. The parasitic *A. m. capensis* workers appear to have no specific advantage that aids them in horizontal transmission. Guards at entrances of *A. m. scutellata* colonies reject non-nestmate *A. m. capensis* workers with the same probability as non-nestmate *A. m. scutellata* workers¹². However, special mechanisms facilitating horizontal transmission are probably unnecessary because entrance guards usually accept a large proportion of non-nestmates attempting to enter^{14,15}, and workers frequently 'drift' among colonies within an apiary¹⁶.

To our knowledge this is the first example of parasitism of eusocial colonies by self-replicating workers. However, this was perhaps foreseen by William Hamilton (ref. 17) who noted that "An ability of females to lay unfertilized eggs which develop into females would open another possible avenue for selfish selection". This parasitism is analogous to cancer in that it is by same-species, self-replicating units that do not respond to normal regulatory processes and proliferate to overwhelm the collective¹⁸. It is also analogous to a

normal infectious disease in that the replicating workers are transmitted horizontally between hosts rather than arising *de novo* within each host colony. □

Methods

Study organisms

All *A. m. scutellata* colonies, eggs and worker bees were from eight colonies moved into Pretoria from a region free of the *A. m. capensis* parasitism in October 2000, just before the study. All *A. m. capensis* eggs and workers were obtained from infected colonies provided by a bee-keeper in the Pretoria area. These colonies showed the typical symptoms of the *A. m. capensis* problem: colony dwindling, multiple eggs per cell, and dark-coloured bees (*A. m. scutellata* workers are more yellow than the parasitic *A. m. capensis* workers). In addition a subsample of our *A. m. capensis* bees ($n = 5$) are genetically the same as parasitic *A. m. capensis* workers from across northern South Africa (P. Kryger, personal communication).

Egg laying and worker policing by *A. m. scutellata*

Following standard methods⁶, each *A. m. scutellata* study colony was queenright and housed in two hive boxes with a queen excluder between the upper and lower box (the queen was in the lower box). One test frame per colony containing thousands of worker cells and 500–900 empty drone cells was placed in the upper box and sandwiched between two frames containing open brood. The drone cells in each frame were checked daily and the locations of any eggs were noted⁶. The five eggs observed (see Table 1) were all removed within 24 h. This procedure shows that both egg laying by workers and worker policing occurs in *A. m. scutellata*.

Egg-removal rates

Following standard methods^{5,9,10} we transferred 600 *A. m. scutellata* queen-laid eggs, 600 *A. m. scutellata* worker-laid eggs and 600 parasitic *A. m. capensis*-laid eggs into worker- and drone-sized cells in three test frames (that is, 20 eggs $\times$ 3 discriminator colonies $\times$ 5 days $\times$ 2 cell sizes per egg source). Test frames were placed into unrelated *A. m. scutellata* discriminator colonies above a queen excluder and sandwiched between two frames containing open brood. Egg removal was quantified by inspecting cells after 2, 6 and 20 h. The difference among all egg sources after 2 h was highly significant ($P < 0.001$). No significant difference between cell type and day or interaction between any of the above parameters was found (three-way analysis of variance (ANOVA)). The data were arcsine-transformed to satisfy the assumption of normality because most observations were close to the limits.

We obtained *A. m. capensis* eggs from infected *A. m. scutellata* colonies. We verified that the eggs were laid by *A. m. capensis* workers and not an *A. m. scutellata* queen as follows: only *A. m. capensis* workers were emerging from cells; there was an irregular pattern of eggs in cells (with multiple eggs in some cells), as is typical in a colony with egg-laying workers; no *A. m. scutellata* queen was seen in any colony despite several hive inspections each.

Received 28 August; accepted 8 November 2001.

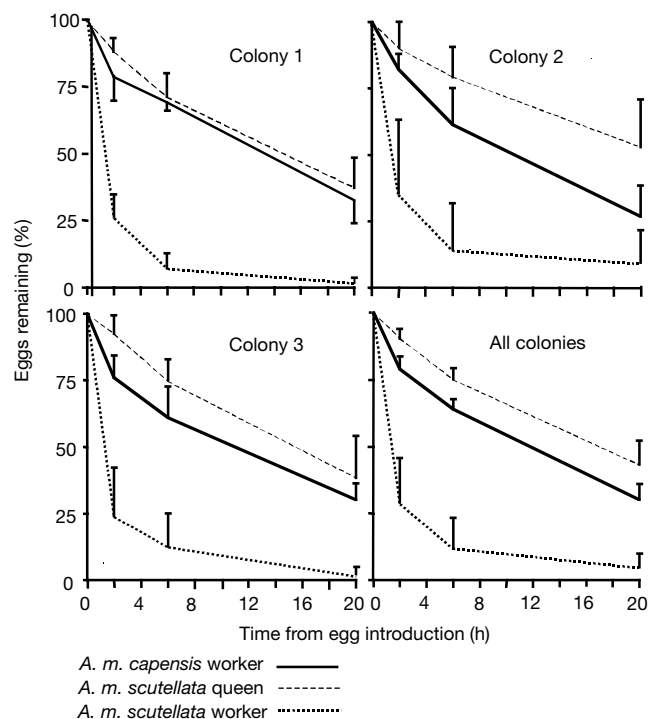


Figure 2 Removal rates of eggs (mean plus s.d.) laid by *A. m. scutellata* queens (dashed line), *A. m. scutellata* workers (dotted line) and parasitic *A. m. capensis* workers (solid line) when introduced into three queenright *A. m. scutellata* discriminator colonies (see Methods for detailed experimental procedure).

- Allsopp, M. H. & Crewe, R. M. The cape honey bee as a Trojan horse rather than the hordes of Genghis Khan. *Am. Bee J.* **133**, 121–123 (1993).
- Verma, S. & Ruttner, F. Cytological analysis of the thelytokous parthenogenesis in the Cape honeybee (*Apis mellifera capensis* Escholtz). *Apidologie* **14**, 41–57 (1983).
- Kryger, P. in *Proc. 13th Entomol. Congr. S. Afrca* (eds Olckers, T. & Brothers, D. J.) 38–39 (Entomological Society of Southern Africa, Pretoria, 2001).
- Ratnieks, F. L. W. Reproductive harmony via mutual policing by workers in eusocial Hymenoptera. *Am. Nat.* **132**, 217–236 (1988).
- Ratnieks, F. L. W. & Visscher, P. K. Worker policing in the honey bee. *Nature* **342**, 796–797 (1989).
- Ratnieks, F. L. W. Egg-laying, egg-removal, and ovary development by workers in queenright honey bee colonies. *Behav. Ecol. Sociobiol.* **32**, 191–198 (1993).
- Barron, A., Oldroyd, B. P. & Ratnieks, F. L. W. Worker reproduction in honey-bees (*Apis*) and the anarchic syndrome: a review. *Behav. Ecol. Sociobiol.* **50**, 199–208 (2001).
- Visscher, P. K. Reproductive conflict in honey bees: a stalemate of worker egg-laying and policing. *Behav. Ecol. Sociobiol.* **39**, 237–244 (1996).
- Ratnieks, F. L. W. Evidence for a queen-produced egg-marking pheromone and its use in worker policing in the honey bee. *J. Apic. Res.* **34**, 31–37 (1995).
- Oldroyd, B. P. & Ratnieks, F. L. W. Anarchistic honey bee workers evade worker policing by laying eggs that have low removal rates. *Behav. Ecol. Sociobiol.* **47**, 268–273 (2000).
- Oldroyd, B. P., Smolenski, A. J., Cornuet, J. M. & Crozier, R. H. Anarchy in the beehive. *Nature* **371**, 749 (1994).
- Beekman, M., Wossler, T. C., Martin, S. J. & Ratnieks, F. L. W. Acceptance of Cape honey bees (*Apis mellifera capensis*) by African guards (*A. m. scutellata*). *Insectes Sociaux* (submitted).
- Martin, S., Wossler, T. C. & Kryger, P. The dynamics of usurpation of *Apis mellifera scutellata* colonies by parasitic *A. m. capensis* workers. *Apidologie* (in the press).
- Downs, S. G. & Ratnieks, F. L. W. Adaptive shifts in acceptance thresholds by guard honey bees. *Behav. Ecol.* **9**, 326–333 (2000).
- Downs, S. G., Ratnieks, F. L. W., Badcock, N. & Mynott, A. Honey bee guards do not use food-derived odours to recognise non-nestmates: a test of the odour convergence hypothesis. *Behav. Ecol.* **12**, 47–50 (2001).
- Neumann, P., Moritz, R. F. A. & Mautz, D. Colony evaluation is not affected by drifting of drone and worker honeybees (*Apis mellifera* L.) at a performance testing apiary. *Apidologie* **31**, 67–79 (2000).
- Hamilton, W. D. The genetical evolution of social behaviour. *J. Theor. Biol.* **7**, 17–52 (1964).
- Nunney, L. in *Levels of Selection in Evolution* (ed. Keller, L.) 238–252 (Princeton Univ. Press, Princeton, 1999).

Acknowledgements

This research was funded by the Natural Environment Research Council (UK) and the 'Social Evolution' Network financed by the EC Training and Mobility of Researchers program. A. Shella provided the *A. m. capensis*-infected colonies.

Correspondence and requests for materials should be addressed to E.L.W.R. (e-mail: F.Ratnieks@Sheffield.ac.uk).

Dynamic coding of behaviourally relevant stimuli in parietal cortex

Louis J. Toth & John A. Assad

Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA

A general function of cerebral cortex is to allow the flexible association of sensory stimuli with specific behaviours. Many neurons in parietal^{1,2}, prefrontal^{3,4} and motor⁵⁻⁷ cortical areas are activated both by particular movements and by sensory cues that trigger these movements, suggesting a role in linking sensation to action. For example, neurons in the lateral intraparietal area (LIP) encode both the location of visual stimuli and the direction of saccadic eye movements^{8,9}. LIP is not believed to encode non-spatial stimulus attributes such as colour^{10,11}. Here we investigated whether LIP would encode colour if colour was behaviourally linked to the eye movement. We trained monkeys to make an eye movement in one of two directions based alternately on the colour or location of a visual cue. When cue colour was relevant for directing eye movement, we found a substantial fraction of LIP neurons selective for cue colour. However, when cue location was relevant, colour selectivity was virtually absent in LIP. These results demonstrate that selectivity of cortical neurons can change as a function of the required behaviour.

Two rhesus monkeys were trained on a delayed saccade paradigm (Fig. 1). Monkeys fixated a spot in the centre of the screen, viewed a peripheral cue for 500 ms, held fixation with no cue present for 1 s, and then directed their eyes from the fixation point to one of two identical targets. The peripheral cue appeared in one of two colours and one of two locations. Two complementary colours were chosen at random for each cell. Of the two locations used, one was centred on the cell's 'response field' (mapped independently using a standard delayed-memory-saccade task, see Methods) and the other location was placed 180° opposite, at the same distance from the fixation point. In interleaved blocks of trials, the rule for determining the appropriate target was switched. One rule was to direct the saccade on the basis of cue colour (for example, 'red means saccade left' and 'green means saccade right'), regardless of the cue location. The alternative rule was to direct the saccade to the cue location ('cue on left means saccade left', 'cue on right means saccade right'), regardless of the cue colour. These two rules are termed 'colour-relevant' and 'location-relevant'. In the colour-relevant task, colour-saccade pairings were reversed every 40 trials (for example, to 'red means saccade right' and 'green means saccade left'), so that selectivity for colour could be analysed independently from selectivity for saccade direction. The colour-relevant task included antisaccades (eye movements away from cue) as well as prosaccades (eye movements towards cue) to require the animals to rely on colour alone to perform the task (Fig. 1). Prosaccade trials in the colour-relevant task matched exactly (in cue colour, cue location and saccade direction) prosaccade trials in the location-relevant task—the only difference was the animal's strategy for choosing. For example, in all trials where a red cue appeared on the left and the

animal correctly made a leftward eye movement, the animal was correct on colour-relevant trials because the cue was red, and on location-relevant trials because the cue was on the left. Thus we were able to determine the effect of the behavioural rule on the stimulus selectivity of the neurons.

Figure 2 shows data recorded from 69 LIP neurons during the location-relevant task. Selectivity for cue location (Fig. 2a, c, e) and colour (Fig. 2b, d, f) was examined in three ways. First, normalized population-response histograms were constructed, averaged across all cells. For each cell, trials were sorted by preferred (black line) and non-preferred (grey line) cue location (Fig. 2a) and colour (Fig. 2b). Because the preferred colour was defined as the colour that gave the greater response in this task, colour selectivity was compared to a chance selectivity level computed by a bootstrap technique (see Methods). Although the population of cells was strongly selective for cue location (Fig. 2a), no colour selectivity above chance was observed (Fig. 2b). Two measures were used to quantify the colour and cue-location selectivity. First, transmitted information^{12,13}, which indicates the reliability of differences in firing rate between two sets of trials, was separately calculated for cue location and colour (Fig. 2c, d). Second, an analysis of variance (ANOVA) was calculated for each unit to evaluate the significance of modulation by cue location or colour ($\alpha = 0.05$; Fig. 2e, f). Both of these analyses were calculated using spike counts in each of five 500-ms epochs of the trial: baseline (B), cue visible (C), early delay (D1), late delay (D2), and perisaccade (S) (see Methods). In the location-relevant task, up to 49% of neurons showed significant selectivity for cue location in any single time period (Fig. 2c, e) while no more than 13% of neurons showed significant selectivity for colour (Fig. 2d, f).

Responses of neurons to corresponding prosaccade trials of the colour-relevant task were analysed with the same methods (Fig. 3). As in the location-relevant task, population responses showed substantial modulation by cue location (Fig. 3a), but unlike in the

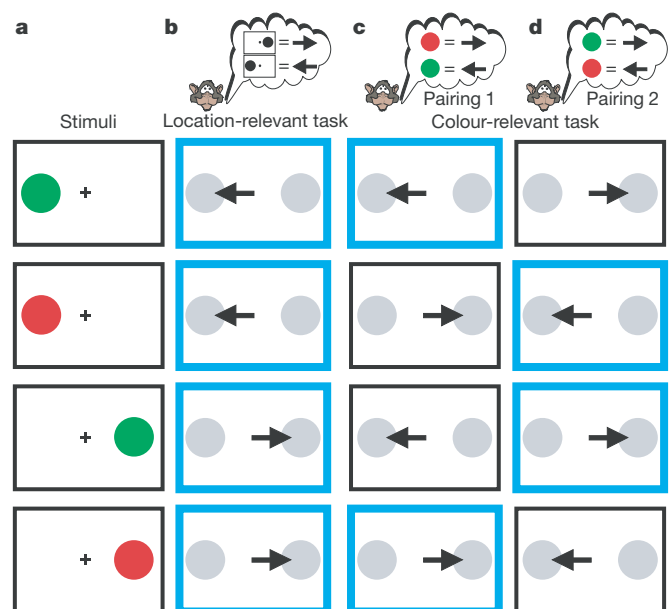


Figure 1 Design of experiment. On each trial, the animal was shown a visual cue (a) and after a delay given two identical saccade targets. For the location-relevant task (b), the location of the cue, regardless of its colour, indicated which target to choose. For the colour-relevant task (c, d) the colour of the cue, regardless of its location, indicated which target to choose. Colour-saccade pairings were alternated in blocks of trials yielding all possible pairings of colour, cue location, and saccade direction. In each row, prosaccade trials (blue boxes) are identical in cue colour, cue location and saccade direction in both colour-relevant and location-relevant tasks. Cartoons indicate the behaviour rules that the trained monkeys applied for each task.

location-relevant task, the population responses were also selective for colour (Fig. 3b). Information analysis and ANOVA confirmed the greater colour selectivity. In the colour-relevant task, up to 53% of neurons (in any single time period) were significantly modulated by cue location (Fig. 3c, e), and up to 29% of neurons showed significant colour modulation (Fig. 3d, f). Across the population, colour information in the location-relevant task did not differ from the baseline level in any period of the task, while in the colour-relevant task colour information was significantly different from the baseline level in cue, early-delay and late-delay periods (Mann–Whitney *U* test, $P < 0.05$). Thus the same population of neurons was selective for cue location in both tasks, but was selective for colour preferentially during the colour-relevant task.

Several features of these cells' colour specificity were notable. First, colour selectivity was not due merely to increased spiking during the colour-relevant task. If the neuronal responses during the location-relevant task were smaller, colour modulation might have been more difficult to detect in that task. Rather, we observed slightly larger average cue responses in the location-relevant task than in the colour-relevant task.

Second, the colour selectivity was independent of the association of a particular colour with the preferred saccade direction. To compare directly the colour-relevant and location-relevant tasks only prosaccade trials were analysed, because those trials were physically identical between the two tasks (Figs 2 and 3). In a separate analysis, antisaccade trials from the colour-relevant task were sorted by the preferred colour judged from prosaccade trials, and averaged across the population (Fig. 4a). On average, the same colour preference was seen. To examine whether individual cells maintained the same colour selectivity on prosaccade and antisac-

cade trials, colour information was calculated for antisaccade trials and 'signed' by whether the preferred colour was the same or opposite from that for prosaccade trials (Fig. 4b). Comparable proportions of neurons were significantly colour-selective for prosaccade and antisaccade trials (C, 29% versus 24%; D1, 25% versus 27%; D2, 21% versus 17%; ANOVA, $P < 0.05$). Across the population, prosaccade and antisaccade trials were much more likely to show the same colour preference than the opposite colour preference (one-tailed *t*-test; $H_0: \mu_{\text{antisaccade}} \leq 0$; $P < 0.0005$ for C, D1 and D2 periods). In particular, for C, D1 and D2 periods no fewer than 80% of cells with significant colour selectivity for antisaccade trials had the same preferred colour for prosaccade trials. These data argue that the colour selectivity was not an artefact of nonspecific changes in firing rate across blocks, because antisaccade trials should then have shown the opposite selectivity for the two colours.

Third, colour modulation was not due to luminance modulation. As a control, the stimulus was assigned a random luminance on each trial, spanning $\sim 25\%$ contrast range around the point where both colours were physically equiluminant as measured with a photometer. This ensured that animals could not use cue luminance to solve the task. Trials were sorted into highest and lowest cue luminance, and transmitted information for high versus low luminance was calculated as in Figs 2 and 3. Across the population, luminance information did not differ from the baseline level in any period of the task (Mann–Whitney *U* test, $P > 0.05$), indicating that the colour selectivity observed was not due to differences in cue luminance.

These experiments show that LIP neurons can be selective for stimulus colour if that colour is relevant for encoding an upcoming eye movement. LIP is part of the dorsal stream 'where' pathway and shows activity related to spatial attention and saccade planning'. LIP projects directly to saccade-generating circuitry via the superior

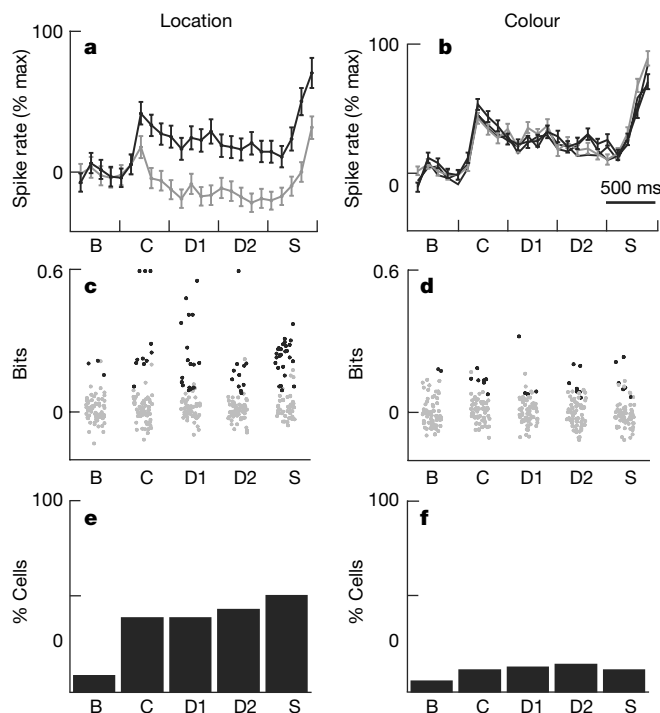


Figure 2 Three analyses of cue-location selectivity and colour selectivity for the location-relevant task. **a, b**, Population-response histograms sorted by cue location (**a**) and by preferred versus null colour (**b**). Error bars are ± 1 standard error of the mean (s.e.m.). Thin lines in **b** indicate the chance level of difference. **c, d**, Transmitted information about cue location (**c**) and colour (**d**) calculated from spike counts over each epoch of the experiment. Each point represents one epoch for one cell. Black points indicate cells with statistically significant colour information above bootstrap (see Methods). For clarity, points are spread horizontally and values over 0.6 are plotted as 0.6. **e, f**, Percentage of cells significant by analysis of variance (ANOVA) for cue location (**e**) and colour (**f**). $n = 69$ cells.

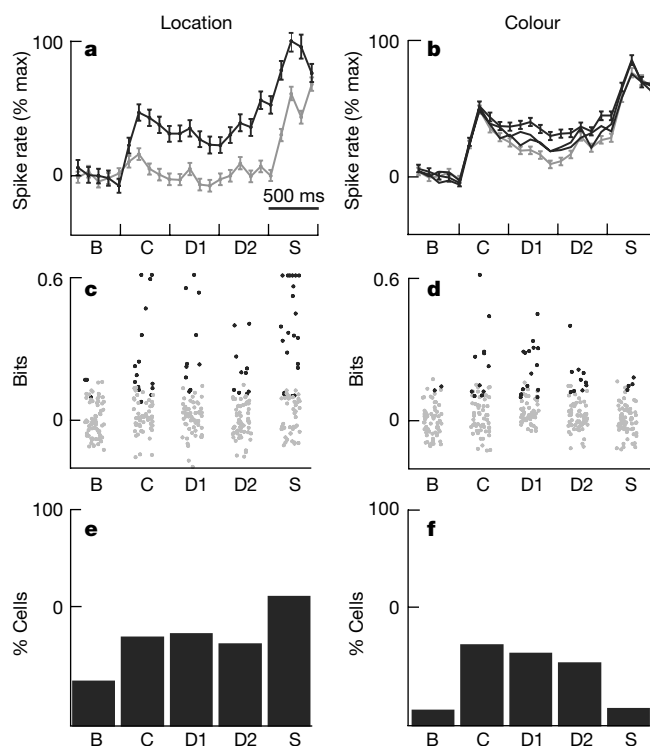


Figure 3 Three analyses of cue-location selectivity and colour selectivity for the colour-relevant task. Conventions as in Fig. 2. **a, b**, Population-response histograms sorted by cue location (**a**) and by preferred versus null colour (**b**). **c, d**, Transmitted information about cue location (**c**) and colour (**d**). **e, f**, Percentage of cells significant by ANOVA for cue location (**e**) and colour (**f**). $n = 118$ cells, except in **c, d**, where n is restricted to the 69 cells shown in Fig. 2 (see Methods) for comparison purposes.

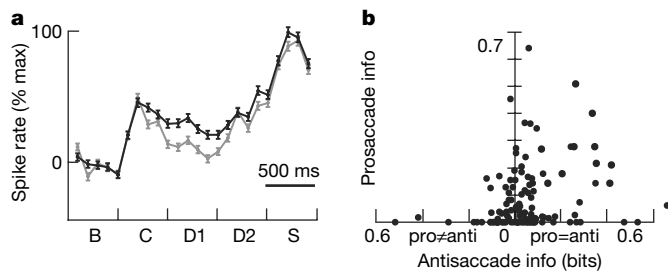


Figure 4 Colour selectivity in antisaccade trials of the colour-relevant task. **a**, Antisaccade trials sorted by the preferred colour (black line) and non-preferred colour (grey line) judged from prosaccade trials. Error bars are ± 1 s.e.m. **b**, Cue-period colour information for each cell, for antisaccade trials (x axis) and prosaccade trials (y axis). Values to right of zero indicate that preferred colour was the same between prosaccade and antisaccade trials, while values to left of zero indicate that preferred colour was opposite in prosaccade and antisaccade trials.

colliculus¹⁴ and FEF¹⁵, and saccades can be elicited from LIP with relatively low levels of electrical stimulation¹⁶. Thus LIP has mainly been thought to specify spatial attributes of visual scenery in a manner suitable for encoding saccadic eye movements⁸. (Shape selectivity has also been reported in LIP, even during passive fixation¹⁷, although this may reflect the inherently spatial nature of shape processing.) In tasks involving a fixed association of specific coloured cues with specific movements, LIP neurons, strictly speaking, can respond selectively to the cues, but this apparent colour selectivity is almost certainly due to selectivity for the associated movement^{18,19}. In our task, rather than having a fixed association between cue colour and saccade direction, the association of colour and movement was constantly reversed, allowing colour selectivity to be determined independently from movement selectivity. These combined results suggest that the timescale or the context of the association may be crucial in determining whether LIP encodes colour information.

Our data support the hypothesis that LIP can directly encode arbitrary stimulus characteristics if they are relevant for guiding eye movements²⁰. LIP may receive colour information via a direct projection from a colour-sensitive visual area such as V4 (ref. 15). Alternatively, it may be that an as yet unidentified area is responsible for identifying the relevant stimulus attribute and passing on its quale to LIP. The prefrontal cortex is a candidate for such an area since activity specific to the switching of a behavioural task has been reported there²¹. LIP may thus receive the end product of attentional gating, as selectivity in LIP appears to be specific for behaviourally relevant features. □

Methods

Two rhesus monkeys (male, ~10 kg) were implanted with a head post, scleral search coil²², and recording chamber. LIP was approached dorsally, guided by magnetic resonance imaging (MRI) (echo time/relaxation time 7.0/14.3, field of view 190 × 190, 256 pixels). In each penetration, the medial and lateral banks of the intraparietal sulcus (IPS) were encountered sequentially. Using a delayed-memory-saccade (DMS) task, neurons with strong, selective delay-period activity indicative of LIP were found in the lateral bank of the caudal IPS (coordinates P2 L10–A2 L13). Neurons were considered to be in LIP if they had selective delay-period activity or, in the same electrode penetration, were located between such cells.

Response fields were mapped with an accuracy of 3° radius and 30° angle relative to the fovea using a DMS task and hand mapping. Subsequent stimuli (coloured cues and white gabor-modulated targets, 3–5° diameter) were placed in the centre of the response field and 180° opposite to it at the same radial distance. A pair of complementary colours was chosen randomly for each cell. Fixation within 1° and saccade within 2.5° of target were required. Colour- and location-relevant tasks were run in alternating blocks containing at least 160 and 80 correct trials respectively. Within colour-relevant blocks, pairings were switched every 40 correct trials to dissociate colour and cue-location information. For each cell, the first task was randomly chosen and tasks were alternated at least A-B-A. The proportion of colour-selective neurons in the colour-relevant task was comparable whether the colour-relevant or location-relevant trials came first (25.3% versus 27.0%,

averaged across C, D1 and D2 periods). Approximately equal numbers of trials were collected from each cell for each of the conditions shown in Fig. 1. Once trained, animals discerned any rule or pairing within a few trials; therefore all trials were analysed. Data were discarded if performance fell below 75%. For the two animals, the overall mean performance ($\pm$ s.e.m.) was $81.8 \pm 2.4\%$ and $88.8 \pm 2.1\%$. Each animal's performance was not significantly different for trials sorted either by preferred/null colour, colour pairing or prosaccade/antisaccade (t -test, $P > 0.05$). Also, for each cell, mean saccade amplitude, peak velocity, and latency were compared between preferred-colour and null-colour trials and between prosaccade and antisaccade trials. No significant differences were found across the population (paired t -test; $P > 0.05$), and fewer than 10% of individual cells showed significant differences for any of the three saccade metrics (t -test, $\alpha = 0.05$). Those few significant cells showed no systematic trend towards larger saccade metrics in prosaccade trials or preferred-colour trials.

Data analysis

For population histograms, each cell's responses were normalized to the baseline and the peak firing rate calculated by averaging across all trial types. The responses of each cell to preferred and null colour and preferred and null location were then averaged across the population, so that percentage modulation of maximal response across all conditions is expressed. For analysing colour selectivity, responses were combined across both locations, and vice versa. Preferred location was determined from the DMS task. Preferred colour was assessed from the delay periods of trials where the cue appeared in the response field. For the colour-relevant task, the preferred colour was chosen from prosaccade trials only. Bootstrap lines were calculated by randomly assigning trials into two groups, labelling the group with the larger average as 'preferred' and generating histograms as above.

Transmitted information and ANOVA were used to assess the effects of colour and cue location independently for spike counts from five 500-ms periods of each trial: baseline (animal fixating), cue, delay (early and late, 500 ms each), and perisaccade (after targets appeared and during the saccade). We used information analysis to ask whether responses in the two conditions were distinguishable from each other (1 bit maximum). Trials were combined across the irrelevant dimension, and each period was analysed independently. Significance was calculated for each cell by a bootstrap procedure: values of the input variable (colour or cue location) were shuffled across trials and the information calculation repeated 1,000 times. Cells were judged to have significant modulation if the information was greater than 95% of the information obtained using randomly shuffled input. For each unit we subtracted the chance-level information obtained from the bootstrap from the actual information value. Information and ANOVA were in good agreement on the significance of modulation of individual cells. Similar results were obtained with bin size halved or doubled.

Initially, 49 cells were tested with only the colour-relevant task. These cells showed qualitatively similar selectivity for colour to the other 69 cells (C, 33% versus 28%; D1, 20% versus 29%; D2, 20% versus 22%) and were found in the same region; they were therefore included in colour-relevant task analyses. Of the 118 total cells tested with the colour-relevant task, colour-selective neurons were found in both animals (animal D: $n = 79$; C, 27%; D1, 19%; D2, 19%; animal R: $n = 39$; C, 32%; D1, 30%; D2, 22%).

Received 15 June; accepted 31 October 2001.

- Newsome, W. T., Wurtz, R. H. & Komatsu, H. Relation of cortical areas MT and MST to pursuit eye movements. II. Differentiation of retinal from extraretinal inputs. *J. Neurophysiol.* **60**, 604–620 (1988).
- Andersen, R. A., Snyder, L. H., Li, C. S. & Stricanne, B. Coordinate transformations in the representation of spatial information. *Curr. Opin. Neurobiol.* **3**, 171–176 (1993).
- Murray, E. A., Bussey, T. J. & Wise, S. P. Role of prefrontal cortex in a network for arbitrary visuomotor mapping. *Exp. Brain Res.* **133**, 114–129 (2000).
- Asaad, W. F., Rainer, G. & Miller, E. K. Neural activity in the primate prefrontal cortex during associative learning. *Neuron* **21**, 1399–1407 (1998).
- Johnson, M. T., Coltz, J. D., Hagen, M. C. & Ebner, T. J. Visuomotor processing as reflected in the directional discharge of premotor and primary motor cortex neurons. *J. Neurophysiol.* **81**, 875–894 (1999).
- Georgopoulos, A. P. Neural aspects of cognitive motor control. *Curr. Opin. Neurobiol.* **10**, 238–241 (2000).
- Bichot, N. P., Schall, J. D. & Thompson, K. G. Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature* **381**, 697–699 (1996).
- Andersen, R. A., Snyder, L. H. & Bradley, D. C. Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annu. Rev. Neurosci.* **20**, 303–330 (1997).
- Colby, C. L. & Goldberg, M. E. Space and attention in parietal cortex. *Annu. Rev. Neurosci.* **22**, 319–349 (1999).
- Ungerleider, L. G. & Mishkin, M. in *Analysis of Visual Behavior* (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W.) 549–586 (MIT Press, Cambridge, Massachusetts, 1982).
- Colby, C. L. & Duhamel, J. R. Spatial representations for action in parietal cortex. *Brain Res. Cogn. Brain Res.* **5**, 105–115 (1996).
- Abramson, N. *Information Theory and Coding* (McGraw-Hill, New York, 1963).
- Eskandar, E. N. & Assad, J. A. Dissociation of visual, motor and predictive signals in parietal cortex during visual guidance. *Nature Neurosci.* **2**, 88–93 (1999).
- Lynch, J. C., Graybiel, A. M. & Lobeck, L. J. The differential projection of two cytoarchitectonic subregions of the inferior parietal lobule of macaque upon the deep layers of the superior colliculus. *J. Comp. Neurol.* **235**, 241–254 (1985).
- Blatt, G. J., Andersen, R. A. & Stoner, G. R. Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area LIP) in the macaque. *J. Comp. Neurol.* **299**, 421–445 (1990).

16. Thier, P. & Andersen, R. A. Electrical microstimulation distinguishes distinct saccade-related areas in the posterior parietal cortex. *J. Neurophysiol.* **80**, 1713–1735 (1998).
17. Sereno, A. B. & Maunsell, J. H. Shape selectivity in primate lateral intraparietal cortex. *Nature* **395**, 500–503 (1998).
18. Zhang, M. & Barash, S. neuronal switching of sensorimotor transformations for antisaccades. *Nature* **408**, 971–975 (2000).
19. Snyder, L. H., Batista, A. P. & Andersen, R. A. Coding of intention in the posterior parietal cortex. *Nature* **386**, 167–170 (1997).
20. Mazzoni, P., Bracewell, R. M., Barash, S. & Andersen, R. A. Spatially tuned auditory responses in area LIP of macaques performing delayed memory saccades to acoustic targets. *J. Neurophysiol.* **75**, 1233–1241 (1996).
21. Asaad, W. F., Rainer, G. & Miller, E. K. Task-specific neural activity in the primate prefrontal cortex. *J. Neurophysiol.* **84**, 451–459 (2000).
22. Judge, S. J., Richmond, B. J. & Chu, F. C. Implantation of magnetic search coils for measurement of eye position: an improved method. *Vision Res.* **20**, 535–538 (1980).

Acknowledgements

We thank E. Eskandar and I. Lee for their help with all aspects of this work and C. Pack and R. Born for reading the manuscript.

Correspondence and requests for materials should be addressed to L.J.T. (e-mail: ljtoth@alum.mit.edu).

Fibulin-5 is an elastin-binding protein essential for elastic fibre development *in vivo*

Hiromi Yanagisawa*, Elaine C. Davis†, Barry C. Starcher‡, Takashi Ouchi§, Masashi Yanagisawa§||, James A. Richardson¶ & Eric N. Olson*

* Department of Molecular Biology, † Department of Cell Biology, ‡ Department of Pathology, § Department of Molecular Genetics, and || Howard Hughes Medical Institute, University of Texas Southwestern Medical Center at Dallas, Texas 75390, USA

¶ Department of Biochemistry, University of Texas Health Center at Tyler, Texas 75708, USA

Extracellular elastic fibres provide mechanical elasticity to tissues and contribute towards the processes of organ remodelling by affecting cell–cell signalling^{1,2}. The formation of elastic fibres requires the assembly and crosslinking of tropoelastin monomers, and organization of the resulting insoluble elastin matrix into functional fibres. The molecules and mechanisms involved in this process are unknown. Fibulin-5 (also known as EVEC/DANCE) is an extracellular matrix protein abundantly expressed in great vessels and cardiac valves during embryogenesis, and in many adult tissues including the aorta, lung, uterus and skin, all of which contain abundant elastic fibres^{3,4}. Here we show that fibulin-5 is a calcium-dependent, elastin-binding protein that localizes to the surface of elastic fibres *in vivo*. *fibulin-5*^{−/−} mice develop marked elastinopathy owing to the disorganization of elastic fibres, with resulting loose skin, vascular abnormalities and emphysematous lung. This phenotype, which resembles the cutis laxa syndrome in humans⁵, reveals a critical function for fibulin-5 as a scaffold protein that organizes and links elastic fibres to cells. This function may be mediated by the RGD motif in fibulin-5, which binds to cell surface integrins, and the Ca²⁺-binding epidermal growth factor (EGF) repeats, which bind elastin.

We inactivated the mouse *fibulin-5* gene by replacing exon 1a, intron 1 and the 5' untranslated region of exon 1b with a promoterless *lacZ* and a neomycin-resistant gene (see Supplementary Information). An absence of fibulin-5 protein in *fibulin-5* homozygous mice was confirmed by western blots of kidney extracts and supernatants

from primary skin fibroblasts and aortic smooth muscle cells (see Supplementary Information). Homozygous mutant mice were viable and appeared indistinguishable from wild-type littermates during the first 4 weeks, except that slight looseness of the skin was noticeable by weaning. By approximately 50 days of age homozygous mutant mice showed lax skin, as demonstrated by rough hair, sagging jowls, senescent appearance and excess folds of abdominal skin. These abnormalities became more pronounced with age (Fig. 1a).

On dissection, we found that the lungs of homozygous mutant mice were expanded and contained dilated alveoli, which were most severe in the peripheral regions (Fig. 1c, arrows). Tortuosity and elongation of the aorta was observed, and the ductus arteriosus was

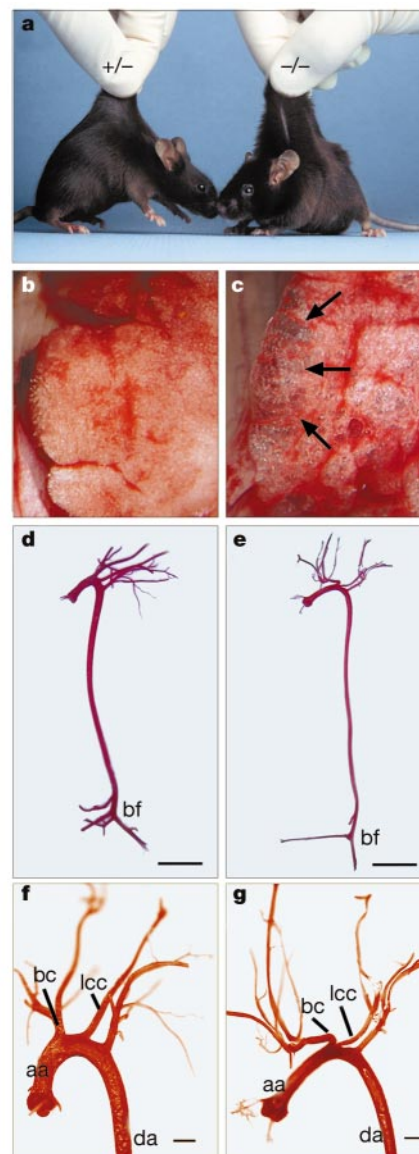


Figure 1 Gross phenotype of *fibulin-5*^{−/−} mice. **a**, The skin of the *fibulin-5*^{−/−} mouse (right) is loose compared with that of the *fibulin-5*^{+/+} littermate (left) at 6 months of age. Note the sagging jowl in the *fibulin-5*^{−/−} mouse. After fingers were released, loose skin folds remained extended in the mutant mouse. **b**, **c**, Right lungs of wild-type (**b**) and *fibulin-5*^{−/−} (**c**) mice at 6 months of age. Note that peripheral portions of the lungs of the *fibulin-5*^{−/−} mouse show multiple air bubbles (arrows). **d**–**g**, Representative casting of the aorta from wild-type (**d**, **f**) and *fibulin-5*^{−/−} (**e**, **g**) mice at 1.5 months of age. Note elongation of the ascending aorta proximal to the aortic arch (**g**). bf, iliac bifurcation of the aorta; aa, ascending aorta; bc, brachiocephalic trunk; lcc, left common carotid artery; da, descending aorta. Scale bars, 0.5 cm (**d**, **e**); 0.1 cm (**f**, **g**).

elongated. To better understand the morphology of the aorta, casting was performed by injection of plastic dye into the left ventricle (Fig. 1d–g). The ascending aorta proximal to the aortic arch was elongated in *fibulin-5*^{-/-} mice, causing the brachiocephalic trunk and left common carotid artery to be juxtaposed (Fig. 1g). The aortic and pulmonary abnormalities were apparent at postnatal day 1 in every *fibulin-5*^{-/-} mouse. Heterozygous *fibulin-5* mice were normal and healthy.

The skin, lung and vascular abnormalities of *fibulin-5*^{-/-} mice were suggestive of a generalized elastinopathy with a possible underlying defect in the development of elastic fibres. Consistent

with this hypothesis, histological examinations of skin from *fibulin-5*^{-/-} mice showed a normal gross appearance (data not shown); however, elastin staining of *en face* sections revealed a marked reduction of the elastic fibre network surrounding hair follicles (Fig. 2a, b). The elastin around the pannicular muscles and the serosa was also significantly reduced and fragmented in homozygous mice (data not shown). In the lung, enlargement of the distal airspaces was already evident at 2 weeks of age in homozygous mice (Fig. 2c, d). This phenotype progressed with age, leading to a severe emphysematous change with the formation of peripheral bullae in adult mice (Fig. 2e, f). Elastin staining showed that elastin at the tips of the secondary septae was fragmented and that the elastic fibres associated with the alveolar walls were short, thickened and disrupted in 2-week-old homozygous mice (Fig. 2h) compared with wild-type littermates (Fig. 2g). Elastin staining of cross-sections of the ascending aorta from wild-type mice showed well organized elastic laminae (Fig. 2i). In contrast, elastic laminae in homozygous mutant mice were fragmented, the defect being more severe towards the adventitia (Fig. 2j). Elastic fibres in muscular arteries and small arteries were also affected (data not shown). Despite the disorganized elastic lamina, there was no indication of aneurysms or of dissections of the medial layers of aortae in *fibulin-5*^{-/-} mice. The morphology of the cardiac valves was normal. Other organs including kidney, intestine, colon, pancreas, liver, uterus, ovary, testis, and the skeleton were normal. Thus, organs most abundant in elastic fibres were affected severely in *fibulin-5*^{-/-} mice.

To examine the ultrastructure of elastic fibres, we performed an electron microscopic analysis of the thoracic aortae of wild-type and *fibulin-5*^{-/-} mice. Electron-dense elastin was seen in both wild-type and homozygous mice (Fig. 3a, b), indicating that tropoelastin was secreted and crosslinked into insoluble elastin. In wild-type mice, continuous elastic laminae were observed with elastin extensions radiating out obliquely from the laminae to the surrounding smooth muscle cells (Fig. 3a). In contrast, the aortae of homozygous

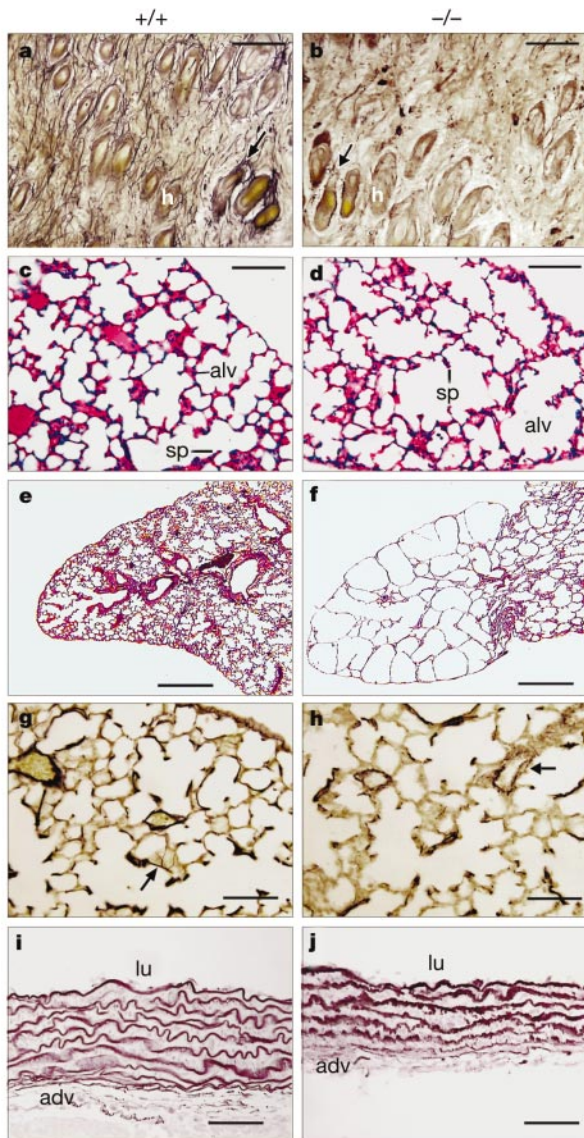


Figure 2 Histological analysis of *fibulin-5*^{-/-} mice. Wild-type (left panels) and *fibulin-5*^{-/-} (right panels) mice are shown. **a, b**, Elastin staining of *en face* sections of the skin. Distinct elastic fibre networks around hair follicles seen in **a** (arrow) are almost absent in **b** (arrow). **h**, hair follicle. **c–f**, Haematoxylin and eosin staining of lung sections from 2-week-old (**c, d**) and 1.5-month-old (**e, f**) mice. The distal airspaces are enlarged in the *fibulin-5*^{-/-} mouse (**d, f**). Note presence of enlarged, disrupted alveoli in **f**. **alv**, alveoli; **sp**, secondary septum. **g, h**, Elastin staining of the lungs from 2-week-old mice. Elastin around the alveolar wall, seen in the lung of the wild-type mouse (**g**, arrow), is fragmented in the lung of the mutant mouse (**h**, arrow). **i, j**, Elastin staining of cross-sections of the ascending aorta. The distinct elastic laminae seen in the wild-type mouse (**i**) are disorganized and fragmented in the *fibulin-5*^{-/-} mutant (**j**). **lu**, lumen; **adv**, adventitia. Scale bars, 100 μ m (**a–d**); 500 μ m (**e, f**); 50 μ m (**g–j**).

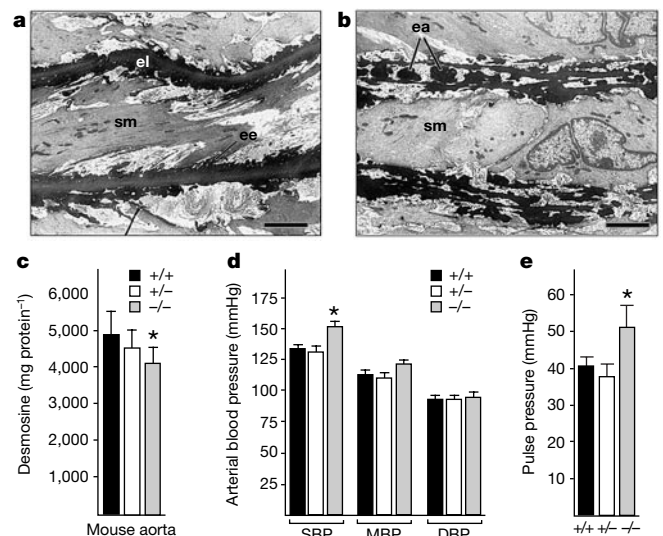


Figure 3 Characterization of aorta in *fibulin-5*^{-/-} mice. **a, b**, Electron micrographs of the thoracic aorta from wild-type (**a**) and *fibulin-5*^{-/-} (**b**) mice. Distinct elastic laminae seen in the wild-type mouse are disrupted in the *fibulin-5*^{-/-} mouse. **ea**, elastin aggregates; **ee**, elastin extension; **el**, elastic lamina; **sm**, smooth muscle cell. Scale bars, 1 μ m. **c**, Measurement of desmosine in aortic arch from 3-month-old wild-type ($n = 8$), *fibulin-5*^{-/-} ($n = 6$) and *fibulin-5*^{-/-} ($n = 11$) mice. **d**, Arterial blood pressure measurement at aortic arch from wild-type ($n = 9$), *fibulin-5*^{+/-} ($n = 8$) and *fibulin-5*^{-/-} ($n = 13$) mice. SBP, systolic blood pressure; MBP, mean blood pressure; DBP, diastolic blood pressure. **e**, Pulse pressure indicated by the difference between SBP and DBP. Asterisk, $P < 0.01$. Values are means $\pm$ s.d. Statistical analysis was carried out using the unpaired student's *t*-test.

mice showed disrupted elastic laminae, which consisted of multiple aggregates of elastin (Fig. 3b). The internal lamina in homozygous mice appeared less affected, being more continuous than the outer laminae (data not shown). No evidence of inflammation was seen throughout the vessel wall in the homozygous mutants, and endothelial cells, smooth muscle cells and adventitial fibroblasts were morphologically normal in mice of both genotypes. In addition, disruption of the elastic laminae was already evident by electron microscopy as early as postnatal day 1 (data not shown). This indicated that the defect seen in adult aortae of homozygous mice was not a result of degradation of intact elastic laminae by activated inflammatory cells, but rather the consequence of an underlying developmental defect in the final organization of the elastic fibres in *fibulin-5*^{-/-} mice.

To estimate elastin content in aortae of wild-type, heterozygous and mutant mice, we ascertained the level of desmosine, an elastin-specific crosslink. As shown in Fig. 3c, there was approximately a 16% decrease in the amount of desmosine in homozygous mutants compared with wild-type mice ($P < 0.01$). There was no significant difference between wild-type and heterozygous mutants. Hydroxy-

proline, an indicator of collagen content, did not differ between wild-type and mutant mice, indicating that collagen fibres were not altered in the aorta of *fibulin-5*^{-/-} mice (data not shown).

We also investigated the physiological effect of disrupted elastic laminae on the haemodynamic properties of the aortic wall. Mean blood pressure and diastolic blood pressure did not differ between wild-type, heterozygous and homozygous mice when measured through catheterization from the aortic arch (Fig. 3d) or the femoral artery (data not shown). However, there was a significant increase in systolic blood pressure measured at the aortic arch in homozygous mice. Thus, pulse pressure in *fibulin-5*^{-/-} mice was significantly increased (Fig. 3e). These results indicate that the pulse-dampening effect of the elastic arteries was disturbed owing to loss of elasticity in the aortic wall in mutant mice. Blood pressure responses to pressor drugs, such as phenylephrine and angiotensin II, and to depressor drugs, such as bradykinin, were normal in homozygous mice (data not shown), indicating that the loss of elasticity did not alter the intrinsic pharmacological properties of smooth muscle cells in *fibulin-5*^{-/-} mice.

To further elucidate the mechanism of fibulin-5 regulation of elastic fibre formation, we performed solid-phase binding assays to assess the biochemical interaction of fibulin-5 with other extracellular matrix proteins. Tropoelastin, fibrillin-1 (amino- and carboxy-terminal halves), laminin, fibronectin, type I collagen and bovine serum albumin (BSA) were used as immobilized protein substrates, and recombinant fibulin-5 was used as a soluble ligand. As shown in Fig. 4a, fibulin-5 bound strongly to tropoelastin and this binding was completely inhibited in the presence of 10 mM EDTA. This suggested that fibulin-5 may use the Ca²⁺-binding EGF-like motifs for binding to tropoelastin. Alternatively, Ca²⁺ may be required for maintaining fibulin-5 in a conformation capable of binding to tropoelastin. No binding between fibulin-5 and laminin or type I collagen was observed (data not shown).

Finally, we determined the precise localization of fibulin-5 in the aorta through the use of immunogold labelling and electron microscopy. In wild-type mice, an intense deposition of gold particles, showing the presence of fibulin-5 protein, was seen along the surface of elastic lamina adjacent to the endothelial cell membrane (Fig. 4b, arrowheads). Immunolabelling was also observed along the elastic laminae in the aortic media (data not shown). No detectable fibulin-5 immunoreactivity was observed in aortae of *fibulin-5*^{-/-} mice (Fig. 4c). These data demonstrated that fibulin-5 co-localizes with elastin, and showed that fibulin-5 is a surface component of elastic fibres *in vivo*.

The formation of mature elastic fibres has been proposed to require the association of elastin with other extracellular matrix proteins, such as fibrillin-1, that aid in assembly and link the fibres to the surface of surrounding cells⁶⁻⁹. However, elastic fibre assembly is unaffected in mice harbouring a mutation in the *fibrillin-1* gene that recapitulates the dominant negative mutation responsible for Marfan syndrome¹⁰, as well as in hypomorphic *fibrillin-1* mutant mice¹¹ or *fibrillin-2*-deficient mice^{12,13}. The marked disorganization of elastic fibres in *fibulin-5*^{-/-} mice described here and by others (T. Nakamura *et al.*, personal communication) reveals a critical role of fibulin-5 as a bridging molecule between elastin and cell surface integrins. Overall, *fibulin-5* mutant mice show the most severe form of elastinopathy to date, thus providing a unique opportunity to study elastogenesis *in vivo*. □

Methods

Generation of mice and Southern blot analysis

Gene targeting was performed as described in Supplementary Information.

Antibody production

Rabbit polyclonal antibody UT65 was raised against rat fibulin-5 polypeptide (CQDINE CEHRNHTCNLQQT; Research Genetics, Cocalico Biologicals), and antibody BSYN1923 was raised against rat fibulin-5 polypeptide (YRGPSYNPYSTSYSGPYAAAPP;

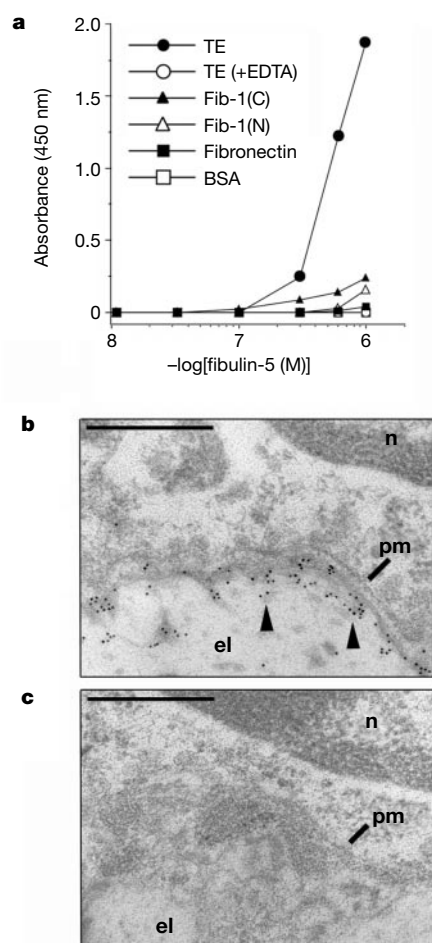


Figure 4 Interaction of fibulin-5 and elastin. **a**, Solid-phase binding assays using recombinant fibulin-5 as a soluble ligand. Immobilized protein substrates are shown in the graph. Absorbance is at 450 nm. Each point represents the average of duplicate (tropoelastin (TE), fibronectin, BSA) or triplicate (fibrillin-1 (Fib-1) N- and C-terminal halves) measurements. Independent experiments were performed three times (TE, fibronectin, BSA) or twice (Fib-1 N- and C-terminal halves). Note that the binding of fibulin-5 to tropoelastin is completely inhibited in the presence of 10 mM EDTA.

b, c, Electron micrograph showing immunolocalization of fibulin-5 in the aorta from wild-type (**b**) and *fibulin-5*^{-/-} (**c**) mice. Note the numerous gold particles on the surface of the internal elastic lamina (arrowheads) of the wild-type aorta (**b**). pm, plasma membrane; n, nucleus of endothelial cell; el, elastic lamina. Scale bars, 0.5 μ m.

Biosynthesis). Purified immunoglobulin- γ (IgG) and affinity-purified antibodies were prepared.

Histology

Mouse tissues were collected and fixed in 10% buffered formalin overnight and embedded in paraffin. Ten-micrometre sections were stained with haematoxylin and eosin for general histological observation, and modified Hart's method was performed for elastin staining.

Aortic casting

Mice were anaesthetized and a 27-gauge winged infusion needle with a syringe was placed in the left ventricle. After incision of the right atrium and abdominal aorta below the renal arteries, PBS containing heparin was infused (0.25 ml min⁻¹) from the left ventricle to flush out blood. The aorta was clamped at the distal end and red Batson's number 17 casting solution (Polysciences) was infused (0.05 ml min⁻¹) until the aorta was filled with the solution. After polymerization, the cast was dissected from the carcass and freed from associated tissues in saturated KOH.

Desmosine analysis

Aortic arches were collected from 3-month-old mice and surrounding tissues were surgically removed. We performed desmosine and hydroxyproline measurements as described¹⁴.

Electron microscopy

Thoracic aortae were dissected from 3-month-old mice after cardiac perfusion with 3% glutaraldehyde in 0.1 M sodium cacodylate. Aortic segments were sequentially stained with osmium tetroxide, tannic acid and uranyl acetate, then dehydrated and embedded in Epon¹⁵. Thin sections (60 nm) were counterstained with uranyl acetate and lead citrate and examined on a Jeol 1200 electron microscope.

Immunolocalization of fibulin-5

Thoracic aortae were dissected from 3-month-old mice after cardiac perfusion with 4% paraformaldehyde in 0.1 M phosphate buffer. Aortic segments were dehydrated in methanol and embedded in Lowicryl K4M (Polysciences). Lowicryl sections on formvar-coated nickel grids were incubated with affinity-purified BSY1923 antibody at a concentration of 0.15 mg ml⁻¹ followed by goat F(ab')₂ anti-rabbit IgG conjugated to 10 nm colloidal gold (Ted Pella). The immunolabelled sections were counterstained with uranyl acetate and lead citrate¹⁶.

Blood pressure measurement

Male mice of 8–12 weeks of age were anaesthetized and a polyethylene catheter (PE-10, Clay Adams) was inserted into the right common carotid artery with the tip at the proximal portion. Mice were then allowed to recover for 24 h before blood pressure measurements. On measurement, the catheter was connected to the pressure transducer (BP-100, CB Sciences), the amplifier (ETH-400, CB Sciences) and a computer-based recorder (PowerLab/8s, CB Sciences), and pulsatile blood pressure was sampled at 200 Hz. Cardiovascular parameters were recorded continuously for more than 2 h under conscious and unrestrained conditions. Systolic, mean and diastolic pressure and heart rate were calculated automatically using the PowerLab system. After measurement of baseline pressure, 0.5 μ g g⁻¹ of vasoactive substances were infused and the haemodynamic effects were examined.

Production of recombinant fibulin-5 and solid-phase binding assay

Recombinant rat fibulin-5 was prepared from a supernatant of a fibulin-5 stable cell line generated with Flp-In-CHO cells (Invitrogen) (see Supplementary Information for details). One microgram of human recombinant tropoelastin without a signal peptide (gift from J. Rosenbloom) was coated on a 96-well maxisorp plate (Nunc) in 10 mM bicarbonate buffer (pH 10) at 4 °C overnight, and 1 μ g of the N- and C-terminal halves of human recombinant fibrillin-1 (rF16 and rF6H (ref. 17), respectively; gift from D. Reinhardt), type I rat collagen (Sigma), rat fibronectin (Sigma), mouse laminin (Sigma) and BSA fraction V (Sigma) were coated in 35 mM carbonate buffer (pH 9.2) at 4 °C overnight. Binding assays with recombinant rat fibulin-5 were performed as described previously¹⁸, with the modification of 2 mM CaCl₂ added in washing buffer. BSYN 1923 (1:500) was used as a primary antibody and horseradish-peroxidase-conjugated goat anti-rabbit IgG (1:3,000, Bio-Rad) was used as secondary antibody. Colour reaction was performed using substrate solution containing tetramethylbenzidine and H₂O₂ (R&D), followed by optic density measurement at 450 nm. Specific binding was calculated by subtracting the background binding of antibodies to immobilized proteins from each reading.

Received 11 September; accepted 13 November 2001.

1. Mecham, R. P. & Davis, E. C. in *Extracellular Matrix Assembly and Structure* (eds Yurchenco, P. D., Birk, D. E. & Mecham, R. P.) 281–314 (Academic, New York, 1994).
2. Li, D. Y. *et al.* Elastin is an essential determinant of arterial morphogenesis. *Nature* **393**, 276–280 (1998).
3. Kowal, R. C., Richardson, J. A., Miano, J. M. & Olson, E. N. EVEC, a novel epidermal growth factor-like repeat-containing protein upregulated in embryonic and diseased adult vasculature. *Circ. Res.* **84**, 1166–1176 (1999).
4. Nakamura, T. *et al.* DANCE, a novel secreted RGD protein expressed in developing, atherosclerotic, and balloon-injured arteries. *J. Biol. Chem.* **274**, 22476–22483 (1999).
5. Milewicz, D. M., Urban, Z. & Boyd, C. Genetic disorders of the elastic fiber system. *Matrix Biol.* **19**, 471–480 (2000).
6. Sasaki, T. *et al.* Tropoelastin binding to fibulins, nidogen-2 and other extracellular matrix proteins. *FEBS Lett.* **460**, 280–284 (1999).

7. Doliana, R. *et al.* EMILIN, a component of the elastic fiber and a new member of the C1q/tumor necrosis factor superfamily of proteins. *J. Biol. Chem.* **274**, 16773–16781 (1999).
8. Trask, B. C. *et al.* Posttranslational modifications of microfibril associated glycoprotein-1. *Biochemistry* **40**, 4372–4380 (2001).
9. Trask, T. M. *et al.* Interaction of tropoelastin with the amino-terminal domains of fibrillin-1 and fibrillin-2 suggests a role for the fibrillins in elastic fiber assembly. *J. Biol. Chem.* **275**, 24400–24406 (2000).
10. Pereira, L. *et al.* Targeting of the gene encoding fibrillin-1 recapitulates the vascular aspect of Marfan syndrome. *Nature Genet.* **17**, 218–222 (1997).
11. Pereira, L. *et al.* Pathogenetic sequence for aneurysm revealed in mice underexpressing fibrillin-1. *Proc. Natl Acad. Sci. USA* **96**, 3819–3823 (1997).
12. Chaudhry, S. S. *et al.* Mutation of the gene encoding fibrillin-2 results in syndactyly in mice. *Hum. Mol. Genet.* **10**, 835–843 (2001).
13. Arteaga-Solis, E. *et al.* Regulation of limb patterning by extracellular microfibrils. *J. Cell Biol.* **154**, 275–281 (2001).
14. Starcher, B. & Conrad, M. A role for neutrophil elastase in the progression of solar elastosis. *Connect. Tissue Res.* **31**, 133–140 (1995).
15. Davis, E. C. Smooth muscle cell to elastic lamina connections in the developing mouse aorta: role in aortic medial organization. *Lab. Invest.* **68**, 89–99 (1993).
16. Davis, E. C. Immunolocalization of microfibril and microfibril-associated proteins in the subendothelial matrix of the developing mouse aorta. *J. Cell Sci.* **107**, 727–736 (1994).
17. Reinhardt, D. P. *et al.* Fibrillin-1: organization in microfibrils and structural properties. *J. Mol. Biol.* **258**, 104–116 (1996).
18. Reinhardt, D. P. *et al.* Fibrillin-1 and fibulin-2 interact and are colocalized in some tissues. *J. Biol. Chem.* **271**, 19489–19496 (1996).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank R. Kowal and B. Abrams for reagents. We thank J. Shelton, J. Stark, S. Yokoyama and A. Rankin for technical assistance, A. Tizenor for graphic assistance, and M. Brown for critical reading of the manuscript. This work was supported in part by the American Heart Association (H.Y.), National Institutes of Health (E.C.D. and E.N.O.) and the D.W. Reynolds Foundation for Clinical Cardiovascular Research (E.N.O.). M.Y. is an investigator for the Howard Hughes Medical Institutes.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.N.O. (e-mail: colson@hamon.swmed.edu) or H.Y. (e-mail: hkyanagisawa@aol.com).

Fibulin-5/DANCE is essential for elastogenesis *in vivo*

Tomoyuki Nakamura*, Pilar Ruiz Lozano*, Yasuhiro Ikeda*, Yoshitaka Iwanaga*†, Aleksander Hinek‡, Susumu Minamisawa*, Ching-Feng Cheng*, Kazuhiro Kobuke§, Nancy Dalton*, Yoshikazu Takada†, Kei Tashiro||, John Ross Jr*, Tasuku Honjo§ & Kenneth R. Chien*

* UCSD-Salk Program in Molecular Medicine and the UCSD Institute of Molecular Medicine, University of California, San Diego, La Jolla, California, 92093, USA

† Vascular Biology, the Scripps Research Institute, La Jolla, California 92037, USA

‡ Division of Cardiovascular Research, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada

§ Department of Medical Chemistry, Graduate School of Medicine, and || Center for Molecular Biology and Genetics, Kyoto University, Kyoto 606-8501, Japan

The elastic fibre system has a principal role in the structure and function of various types of organs that require elasticity, such as large arteries, lung and skin^{1,2}. Although elastic fibres are known to be composed of microfibril proteins (for example, fibrillins and latent transforming growth factor (TGF)- β -binding proteins) and polymerized elastin, the mechanism of their assembly and development is not well understood. Here we report that fibulin-5 (also known as DANCE), a recently discovered integrin ligand³, is an

essential determinant of elastic fibre organization. *fibulin-5*^{-/-} mice generated by gene targeting exhibit a severely disorganized elastic fibre system throughout the body. *fibulin-5*^{-/-} mice survive to adulthood, but have a tortuous aorta with loss of compliance, severe emphysema, and loose skin (cutis laxa). These tissues contain fragmented elastin without an increase of elastase activity, indicating defective development of elastic fibres. Fibulin-5 interacts directly with elastic fibres *in vitro*, and serves as a ligand for cell surface integrins $\alpha\beta 3$, $\alpha\beta 5$ and $\alpha 9\beta 1$ through its amino-terminal domain. Thus, fibulin-5 may provide anchorage of elastic fibres to cells, thereby acting to stabilize and organize elastic fibres in the skin, lung and vasculature.

Fibulin-5, a secreted protein of 448 amino acids and a relative molecular mass of 66,000 (M_r 66K), is strongly expressed by developing arteries³. Fibulin-5 promotes endothelial cell attachment *in vitro* through binding of cell surface integrins to a fibulin-5-coated surface. To study the role of fibulin-5 *in vivo*, we generated *fibulin-5*^{-/-} mice by gene targeting. The first 4 out of 11 exons were deleted and substituted by a β -galactosidase gene (*lacZ*) and neomycin resistance gene driven by the pgk promoter (Supplementary Information Fig. 1a, b). Mice derived from two independent, homologous recombinant embryonic stem cell clones were bred in a C57BL/6 background, and F₂ generation *fibulin-5*^{+/-} animals were mated to generate *fibulin-5*^{-/-} animals. Northern blot analysis at 0.5 days post partum (P) using the full-length fibulin-5 complementary DNA as a probe showed a complete absence of fibulin-5 messenger RNA in *fibulin-5*^{-/-} mice, confirming that the recombinant allele is a null allele (Supplementary Information Fig. 1c). *In situ* hybridization revealed fibulin-5 mRNA expression in the aorta, pulmonary artery and lungs of *fibulin-5*^{+/-} newborn animals (Fig. 1a), whereas expression was not detected in *fibulin-5*^{-/-} littermate animals (Fig. 1b).

fibulin-5^{-/-} mice were born with an expected mendelian frequency, and grew without lethality to at least 6 months of age; however, they had loose skin hanging in folds, a short nose, and large cheeks (Fig. 2a). By dissection, it was also obvious that all *fibulin-5*^{-/-} mice had severe emphysema and extended and tortuous arteries. Aortography revealed marked aortic tortuosity in living *fibulin-5*^{-/-} animals (Fig. 2b, c). We found significant loss of compliance in *fibulin-5*^{-/-} aortae by measuring the diameter of aortic explants while continuously changing the intraluminal pressure (Fig. 2d). These tortuous aortae with loss of compliance are similar to human aged arteries⁴. Histological observation of the lung tissue revealed disruption of alveolar structure in adult *fibulin-5*^{-/-} lungs (Fig. 3a, b). The disruption of distal airways is already found in the lung from P3 neonatal *fibulin-5*^{-/-} mice (Fig. 3c, d), suggesting a developmental defect. Hearts of *fibulin-5*^{-/-} mice showed variable severity of right ventricular enlargement and right-sided

heart failure (Fig. 3e, f), presumably caused by severe emphysema and subsequent pulmonary hypertension, a pathophysiological state known as cor pulmonale⁵. Echocardiography⁶ revealed that some hearts of *fibulin-5*^{-/-} mice showed paradoxical movement of the interventricular septum, which is a sign of severe right ventricular overload⁷ (Fig. 3g, h).

Emphysema in *fibulin-5*^{-/-} mice suggests that these mice may have problems associated with elastic fibres in the lung, because the

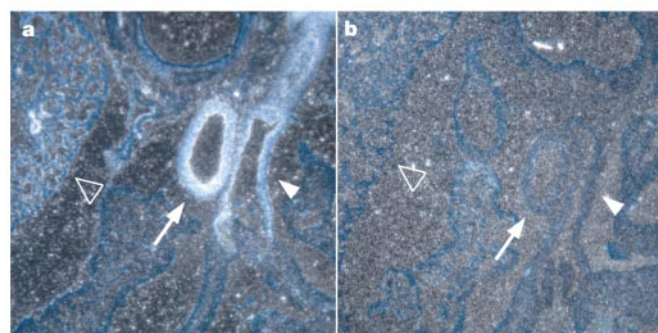


Figure 1 *In situ* hybridization of P0.5 *fibulin-5*^{+/-} (a) and *fibulin-5*^{-/-} (b) transverse sections at the level of the cardiac outflow tract, probed with ³⁵S-labelled fibulin-5 riboprobes. Aortae (arrow), pulmonary arteries (filled arrowhead), and lungs (open arrowhead) are indicated.

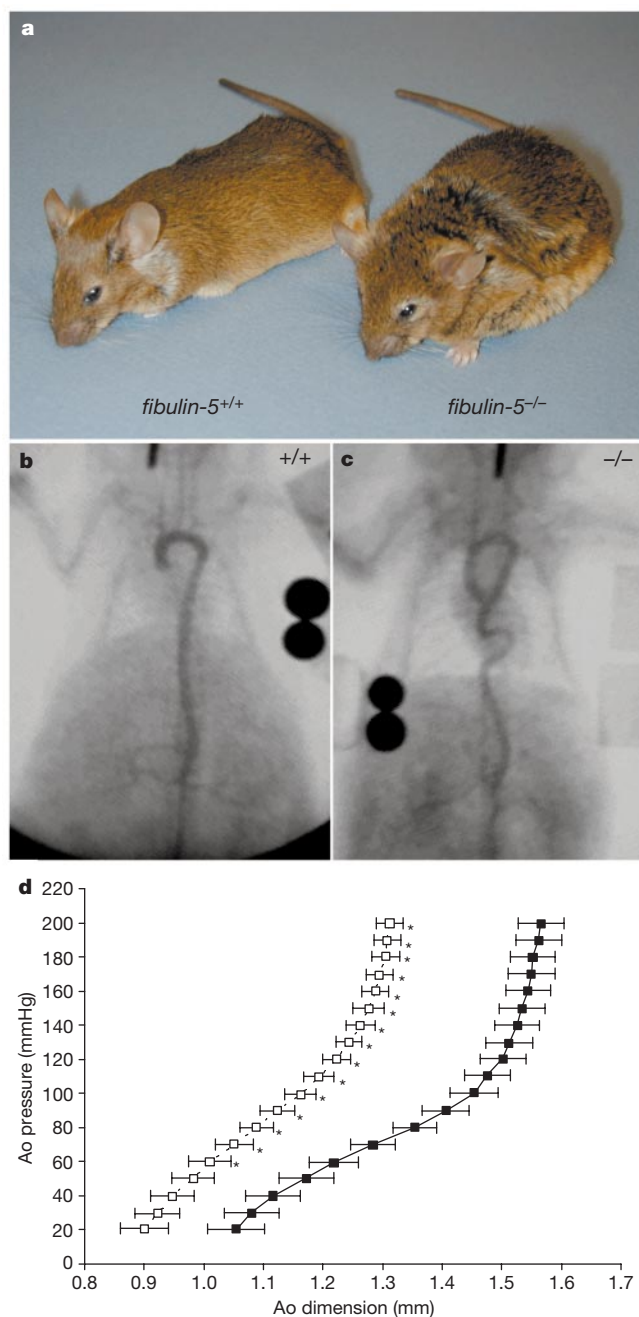


Figure 2 Gross phenotypes and physiological studies of *fibulin-5*^{-/-} mice. **a**, Appearance of *fibulin-5*^{+/-} and *fibulin-5*^{-/-} mice. Loose and wrinkled skin in the *fibulin-5*^{-/-} mouse is noticeable. **b, c**, Aortography of *fibulin-5*^{+/-} (b) and *fibulin-5*^{-/-} (c) mice. Contrast medium was injected at the supravalvular region of the ascending aortae to visualize the lumen of aortae and branches. **d**, Pressure-dimension relationship of the aorta explant. Aortic diameter was measured while continuously changing the intraluminal pressure. *fibulin-5*^{-/-} aortae (open squares) were significantly less extensible than *fibulin-5*^{+/-} aortae (filled squares), even in the range of physiological pressure. Asterisk, $P < 0.01$, compared with *fibulin-5*^{+/-}. Ao, aortic. Error bars, s.e.m.

loss of elastin is associated with human emphysema, and enhancement of elastolysis can cause emphysema^{8,9}. The tortuous artery and loose skin phenotypes also imply loss of elasticity in these organs in *fibulin-5*^{-/-} mice^{10,11}. Thus, we stained tissue sections with elastin van Gieson staining, which specifically stains elastic fibres in black. We found thin and fragmented elastic fibres in skin of *fibulin-5*^{-/-} mice, instead of the thick and long elastic fibres found in skin of *fibulin-5*^{+/+} mice (Fig. 4a, b). With loose skin accompanied by loss of organized elastic fibres, *fibulin-5*^{-/-} mice serve as a model for both cutis laxa, a human disease of loose skin with loss of elastic fibre¹¹, and aged skin, which exhibits less elasticity⁴. There have been no animal models of cutis laxa thus far, although some cases of human cutis laxa are caused by mutations in the Cu²⁺-transporting ATPase gene or the elastin gene¹¹. In descending aortae of *fibulin-5*^{-/-} mice, elastic lamellae were totally disorganized (Fig. 4c, d), with the exception of the innermost layer, which was relatively well preserved. This raises the possibility that endothelial cells might secrete a factor that can compensate for the loss of fibulin-5. Elastic lamellae in the ascending aortae of *fibulin-5*^{-/-} mice were relatively better preserved than in the descending aortae, but still had rough surfaces and were poorly organized compared with *fibulin-5*^{+/+} mice (data not shown). Electron microscopy of aortic sections revealed

the fine structure of the elastic lamellae of the aortae from these mice (Fig. 4e, f, g, h). In *fibulin-5*^{-/-} mice, lumps of polymerized elastin were scattered irregularly, and were not integrated into the dense and organized lamellar structure, as seen in aortae of *fibulin-5*^{+/+} mice (Fig. 4e, f).

In neonates, the amount of polymerized elastin in *fibulin-5*^{-/-} mice was apparently lower than normal (Fig. 4g, h), again suggesting that development of elastic fibre was affected by deletion of the fibulin-5 gene. However, it was still possible that elastic fibre deficiency in *fibulin-5*^{-/-} mice could have resulted from increased elastolytic activity. To test this possibility, we compared the elastolytic activity of tissues and cells from *fibulin-5*^{-/-} and *fibulin-5*^{+/+} mice. Tissue lysates from lung, aorta and skin of *fibulin-5*^{+/+} and *fibulin-5*^{-/-} mice showed no difference in matrix metalloprotease activity by zymography using elastin, gelatin and casein as substrates (data not shown). Assay of elastase activity of skin fibroblasts cultured from *fibulin-5*^{+/+} and *fibulin-5*^{-/-} mice using radiolabelled insoluble elastin also revealed no difference in elastase activity (data not shown), confirming the proposal that fibulin-5 is necessary for

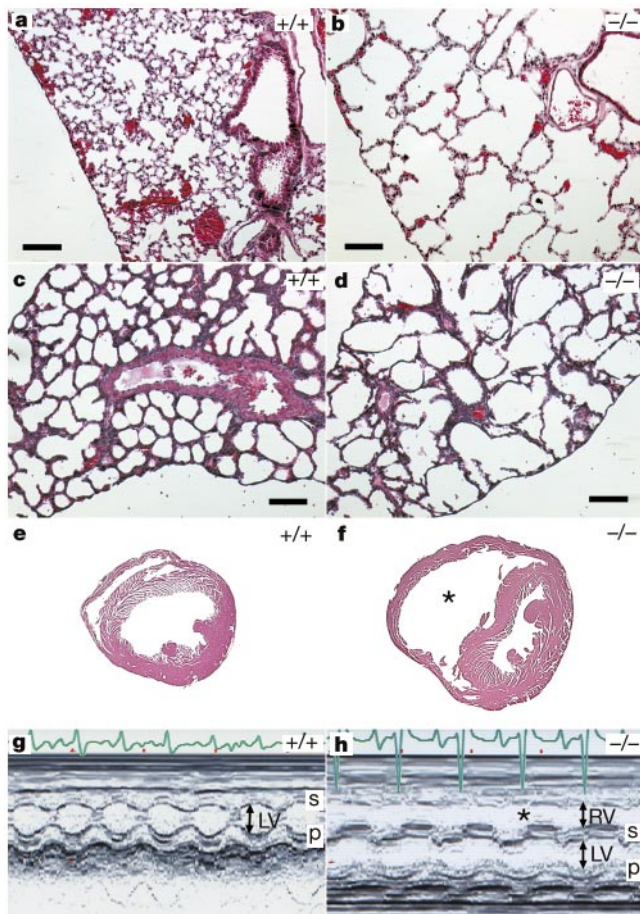


Figure 3 Emphysema and secondary cardiac phenotype in *fibulin-5*^{-/-} mice. **a–d**, Haematoxylin and eosin stained sections of lungs from *fibulin-5*^{+/+} (**a, c**) and *fibulin-5*^{-/-} (**b, d**) mice at 3 months of age (**a, b**) and P3 neonatal stage (**c, d**). Scale bars, 100 μ m. **e, f**, Cross-sections of heart from *fibulin-5*^{+/+} (**e**) and *fibulin-5*^{-/-} (**f**) mice, exhibiting markedly enlarged right ventricle (asterisk) in the *fibulin-5*^{-/-} mouse. Magnification 5.5 $\times$. **g, h**, M-mode echocardiography showing enlarged right ventricle (asterisk) and parallel movement of interventricular septum (s) and posterior wall (p), that is, paradoxical movement in live *fibulin-5*^{-/-} mice (**h**), which is not normally seen in *fibulin-5*^{+/+} mice (**g**). RV, right ventricle; LV, left ventricle.

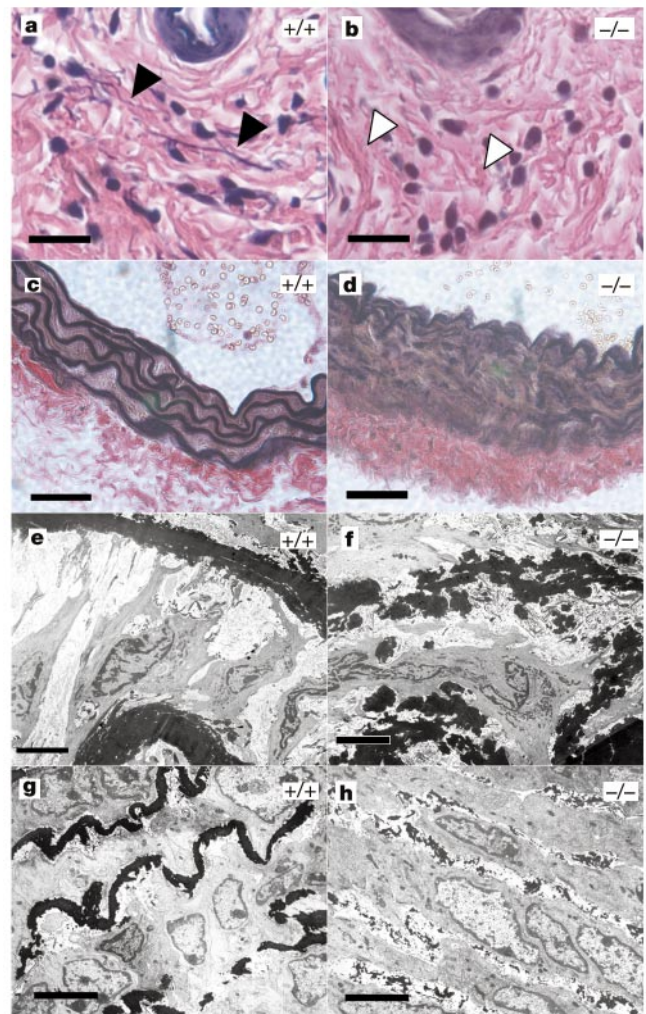


Figure 4 Fine structure of elastic fibres in tissues of *fibulin-5*^{+/+} and *fibulin-5*^{-/-} mice. **a, b**, Elastin van Gieson staining of skin sections. Thick and long elastic fibres (filled arrowheads) were prominent in skins of *fibulin-5*^{+/+} mice (**a**), whereas elastic fibres in *fibulin-5*^{-/-} mice (**b**) were very thin and fragmented (open arrowheads). Scale bars, 20 μ m. **c, d**, Elastin van Gieson staining of transverse sections of thoracic aorta medium from *fibulin-5*^{+/+} (**c**) and *fibulin-5*^{-/-} mice (**d**). Scale bars, 40 μ m. **e–h**, Transmission electron microscopy of adult (**e, f**) and P2 neonate (**g, h**) medial aortic sections of *fibulin-5*^{+/+} (**e, g**) and *fibulin-5*^{-/-} mice (**f, h**). Polymerized, amorphous elastin is stained black with tannic acid. Scale bars, 4 μ m.

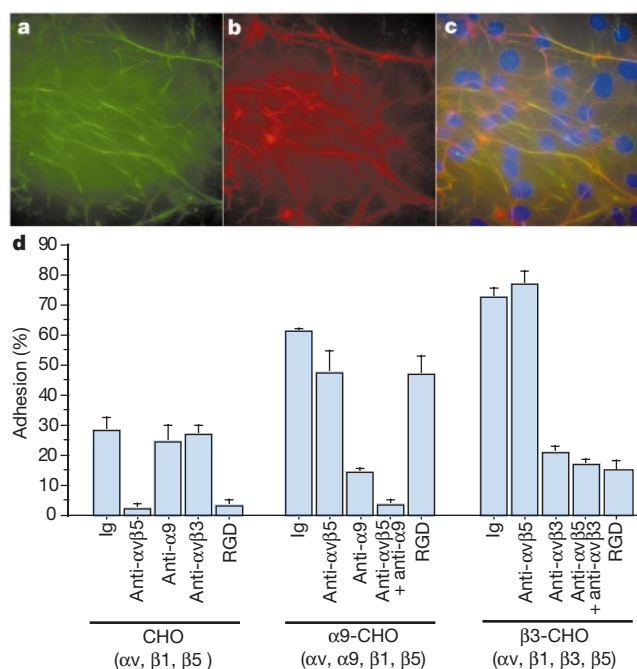


Figure 5 Fibulin-5 binding to elastic fibres and to cell surface integrins. **a–c**, Normal mouse skin fibroblasts were cultured with recombinant Flag-tagged fibulin-5 protein. Cultures were stained with anti-Flag (**a**) and anti-mouse tropoelastin (**b**). **c** is a superimposition of **a** and **b**, together with DAPI nuclear staining showing fibulin-5 localization on elastic fibres. **d**, Cell attachment assay onto recombinant N-terminal,

fibulin-5-domain-coated plates using CHO cells and CHO cell lines stably transfected with integrin α_9 or β_3 chains. The integrin chains present on each cell line are shown in parentheses. A cell attachment assay with full-length human fibulin-5 recombinant protein showed similar integrin specificity (data not shown). Ig, immunoglobulin. Error bars, s.e.m.

the development of elastic fibre, but not for inhibiting elastase activity.

To investigate the possible mechanism by which fibulin-5 mediates development of elastic fibre, we studied whether fibulin-5 interacts with elastic fibres. Mouse skin fibroblasts cultured *in vitro* develop a meshwork of elastic fibres¹¹ (Fig. 5b). Adding Flag-tagged recombinant fibulin-5 to the culture medium resulted in a similar meshwork localization of recombinant fibulin-5 protein (Fig. 5a). Double staining of anti-tropoelastin and anti-Flag antibodies showed co-localization of fibulin-5 and elastin (Fig. 5c), indicating that fibulin-5 could interact directly with elastic fibres. Skin fibroblast cultures of *fibulin-5*^{-/-} mice developed an elastin meshwork to the same extent as cultures from *fibulin-5*^{+/+} mice (data not shown), suggesting that fibulin-5 is not necessary for the initial stages of elastic fibre synthesis, including microfibril assembly and tropoelastin recruitment. Therefore, we propose that fibulin-5 has a crucial role in the maturation of elastic fibres by interacting directly with elastic fibres to promote the higher order organization of thick elastic fibres or elastin lamellae.

As shown by our previous data, fibulin-5 can promote cell attachment through cell surface integrins³. As cell–matrix interactions could also affect tissue organization¹³, we studied the integrin-chain specificity of fibulin-5 attachment by cell attachment assays using the N-terminal domain of fibulin-5 protein and Chinese hamster ovary (CHO) cell stable lines overexpressing various integrin chains (Fig. 5d). Results revealed specific attachment of N-terminal fibulin-5 and integrins $\alpha_v\beta_3$, $\alpha_v\beta_5$ and $\alpha_9\beta_1$, which are known to be expressed by vascular cells^{14,15}. Attachment was inhibited by a blocking antibody to each integrin. Attachment to $\alpha_v\beta_3$ and $\alpha_v\beta_5$ integrins was also inhibited by synthetic RGD-containing peptide, but attachment to $\alpha_9\beta_1$ integrin was RGD-independent, as reported for the attachment of $\alpha_9\beta_1$ integrin and other extracellular matrix proteins^{16,17}. By interacting with elastic fibres and with cell surface integrins, fibulin-5 may stabilize the

attachment of cells to elastic fibres, which could contribute to the organization of nascent elastic fibres during vascular development.

In addition to its structural function, elastin has an anti-proliferative effect on vascular smooth muscle cells, and elastin null mice die perinatally owing to vascular occlusion¹⁸. In *fibulin-5*^{-/-} mice there is no vascular narrowing, reflecting the existence of polymerized elastin in the tissue. Our results indicate that the absence of fibulin-5 does not alter the anti-proliferative effect of elastin, but rather perturbs elastic fibre architecture. Aside from elastin itself, no other single protein has been shown to be essential for elastogenesis and the maturation of elastic fibres in multiple *in vivo* organ systems¹⁹. The loss of elastic fibre structure is associated with several aspects of the ageing phenotype, including the loss of skin elasticity, a marked decrease in arterial compliance, and the onset of a form of chronic lung disease (emphysema)^{20,21}. *fibulin-5*^{-/-} mice now provide an inroad for the dissection of pathways that lead to elastic lamellae formation and related diseases of elastin deficiency. □

Methods

Northern blot analysis and *in situ* hybridization

Total RNA was extracted from the aortae and hearts of neonatal mice using RNAzol (Tel-Test). Ten micrograms of RNA in each lane was separated on a 1% denaturing agarose gel, transferred to an Optitran filter (Schreiber and Schwel), and hybridized with a ³²P-labelled full-length mouse fibulin-5 cDNA. For *in situ* hybridization, paraffin-embedded cross-sections were hybridized with ³⁵S-labelled antisense and sense fibulin-5 complementary RNA probes. Synthesis of the probes, hybridization and washes were performed as described³. Sections were exposed to emulsion for two weeks and analysed with dark-field microscopy.

Aortography and aorta extensibility measurements

For aortography, mice were anaesthetized, ventilated and cannulated from the right carotid arteries. X-ray contrast medium (Mallinckrodt) was injected above the aortic valve, and the real-time image was acquired through posterior–anterior view by an X-ray video capturing system²². For aorta extensibility, the excised aorta was placed in a PBS bath on an inverted microscope connected to a CCD (charge-coupled device) camera, allowing continuous measurement of the dimensions of the outer thoracic aorta. Intra-aortic

chamber pressure was perturbed by syringe from 40 to 200 mmHg, and the external dimensions of the thoracic aorta were obtained simultaneously through a Powerlab data acquisition system (Data Science Instruments). Statistical comparisons of aorta pressure–dimension curves were done by repeated measure of analysis of variance, followed by a Student–Newman–Keuls test. A *P* value of less than 0.01 was considered to be significant.

Histology and electron microscopy

For histology with light microscopy, tissues fixed with 4% paraformaldehyde were embedded in paraffin, sectioned at 8 μm , and stained with haematoxylin and eosin and elastin van Gieson (Sigma). For transmission electron microscopy, tissues were fixed in 4% paraformaldehyde, 2.5% glutaraldehyde, 0.1% tannic acid in 0.1 M cacodylate buffer. Sections were then stained with uranyl acetate.

Cell culture and immunostaining

Normal skin fibroblasts from neonatal mice were cultured in DMEM with 10% fetal bovine serum (FBS) for 10 days at an initial density of $1 \times 10^5 \text{ cm}^{-2}$. We added recombinant Flag-tagged fibulin-5 (ref. 3) to the medium at a concentration of 20 $\mu\text{g ml}^{-1}$. Deposited elastin and recombinant fibulin-5 were detected with anti-mouse tropoelastin antibody (1/30, Elastin Products Company) and anti-Flag M2 antibody (1/100, Sigma), respectively. Rhodamine-anti-rabbit immunoglobulin- γ (IgG) antibody (1/100, Santa Cruz) was used as a secondary antibody for tropoelastin, and fluorescein isothiocyanate (FITC)-anti-mouse IgG antibody (1/100, Santa Cruz) was used as a secondary antibody for Flag. Stained cells were mounted with Vectashield with 4,6-diamidino-2-phenylindole (DAPI) (Vector). Microscope photomicrographs of the same visual fields were superimposed with Photoshop (Adobe).

Cell adhesion assays

The N-terminal domain of human fibulin-5 (amino acids 26–68) glutathione S-transferase (GST) fusion protein was constructed with pGEX-2T plasmid (Amersham Pharmacia) and produced with a bacterial expression system. Cell attachment assays using CHO cells and CHO lines stably transfected with various integrin chains were performed as described^{17,23}. The plates were coated with N-terminal fibulin-5–GST fusion at the concentration of 2 $\mu\text{g ml}^{-1}$. As functional blocking antibodies for integrins, Y9A2 (anti-human α_9 ; Chemicon), LM609 (anti-human $\alpha\text{v}\beta_3$, a gift from D. A. Chersesh) and P1F6 (anti-human $\alpha\text{v}\beta_5$; Chemicon) were used at the concentration of 10 $\mu\text{g ml}^{-1}$ and GRGDSP peptide (Advanced Chem Tech) was used at the concentration of 10 μM .

See Supplementary Information for detailed methods of the construction of the targeting vector and generation of mice.

Received 11 September; accepted 12 November 2001.

- Debelle, L. & Tamburro, A. M. Elastin: molecular description and function. *Int. J. Biochem. Cell Biol.* **31**, 261–272 (1999).
- Rosenbloom, J., Abrams, W. R. & Mecham, R. Extracellular matrix 4: the elastic fiber. *FASEB J.* **7**, 1208–1218 (1993).
- Nakamura, T. *et al.* DANCE, a novel secreted RGD protein expressed in developing, atherosclerotic, and balloon-injured arteries. *J. Biol. Chem.* **274**, 22476–22483 (1999).
- Fazio, M. J., Olsen, D. R. & Uitto, J. J. Skin aging: lessons from cutis laxa and elastoderma. *Cutis* **43**, 437–444 (1989).
- Rapaport, E. in *Textbook of Respiratory Medicine* (eds Murray, J. F. & Nadel, J. A.) 1631–1648 (Saunders, Philadelphia, 2000).
- Minamisawa, S. *et al.* Chronic phospholamban-sarcoplasmic reticulum calcium ATPase interaction is the critical calcium cycling defect in dilated cardiomyopathy. *Cell* **99**, 313–322 (1999).
- Feigenbaum, H. *Ecocardiography* (Lippincott Williams & Wilkins, Baltimore, 1995).
- Piquette, C. A., Rennard, S. I. & Snider, G. L. in *Textbook of Respiratory Medicine* (eds Murray, J. F. & Nadel, J. A.) 1187–1245 (Saunders, Philadelphia, 2000).
- Senior, R. M. & Shapiro, S. D. in *Pulmonary Diseases and Disorders* (ed. Fishman, A. P.) 659–681 (McGraw-Hill, New York, 1998).
- Dobrin, P. B., Schwarcz, T. H. & Mrkvicka, R. Longitudinal retractive force in pressurized dog and human arteries. *J. Surg. Res.* **48**, 116–120 (1990).
- Milewicz, D. M., Urban, Z. & Boyd, C. Genetic disorders of the elastic fiber system. *Matrix Biol.* **19**, 471–480 (2000).
- Robb, B. W., Wachi, H., Schaub, T., Mecham, R. P. & Davis, E. C. Characterization of an *in vitro* model of elastic fiber assembly. *Mol. Biol. Cell* **10**, 3595–3605 (1999).
- Robert, L. Mechanisms of aging of the extracellular matrix: role of the elastin-laminin receptor. *Gerontology* **44**, 307–317 (1998).
- Taooka, Y., Chen, J., Yednock, T. & Sheppard, D. The integrin $\alpha\text{v}\beta_1$ mediates adhesion to activated endothelial cells and transendothelial neutrophil migration through interaction with vascular cell adhesion molecule-1. *J. Cell Biol.* **145**, 413–420 (1999).
- Shattil, S. J. & Ginsberg, M. H. Integrin signaling in vascular biology. *J. Clin. Invest.* **100**, S91–S95 (1997).
- Yokosaki, Y. *et al.* The integrin $\alpha\text{v}\beta_1$ mediates cell attachment to a non-RGD site in the third fibronectin type III repeat of tenascin. *J. Biol. Chem.* **269**, 26691–26696 (1994).
- Eto, K. *et al.* RGD-independent binding of integrin $\alpha\text{v}\beta_1$ to the ADAM-12 and -15 disintegrin domains mediates cell–cell interaction. *J. Biol. Chem.* **275**, 34922–34930 (2000).
- Li, D. Y. *et al.* Elastin is an essential determinant of arterial morphogenesis. *Nature* **393**, 276–280 (1998).
- Dietz, H. C. & Mecham, R. P. Mouse models of genetic diseases resulting from mutations in elastic fiber proteins. *Matrix Biol.* **19**, 481–488 (2000).
- Bailey, A. J. Molecular mechanisms of ageing in connective tissues. *Mech. Ageing Dev.* **122**, 735–755 (2001).

- Labat-Robert, J. & Robert, L. Aging of the extracellular matrix and its pathology. *Exp. Gerontol.* **23**, 5–18 (1988).
- Pashmforoush, M. *et al.* Adult mice deficient in actinin-associated LIM-domain protein reveal a developmental pathway for right ventricular cardiomyopathy. *Nature Med.* **7**, 591–597 (2001).
- Zhang, X. P., Kamata, T., Yokoyama, K., Puzon-McLaughlin, W. & Takada, Y. Specific interaction of the recombinant disintegrin-like domain of MDC-15 (metargidin, ADAM-15) with integrin $\alpha\text{v}\beta_3$. *J. Biol. Chem.* **273**, 7345–7350 (1998).

Supplementary Information accompanies the paper on Nature's website (e-mail: <http://www.nature.com>).

Acknowledgements

We thank Y. Gu, J. Anderson, T. Mikuni, N. Tomikawa and C. Suzuki for technical assistance; P. L. Haywood-Reid for electron microscopy; M. Hoshijima and H. Yasukawa for advice on experimental design; and S. Evans and J. Chen for suggestions in preparing the manuscript. This study was supported by the National Institutes of Health, the Jean LeDucq Foundation, and the American Heart Association Endowed Chair to K.R.C. T.N. was supported by the American Heart Association Western Affiliate Postdoctoral Fellowship.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.R.C. (e-mail: kchien@ucsd.edu).

Stimulated platelets use serotonin to enhance their retention of procoagulant proteins on the cell surface

George L. Dale*, Paul Friese*, Peter Batar*, Stephen F. Hamilton†, Guy L. Reed‡, Kenneth W. Jackson*, Kenneth J. Clemetson§ & Lorenzo Alberio§

* W. K. Warren Medical Research Institute and Department of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104, USA

† College of Pharmacy, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104, USA

‡ Cardiovascular Biology Laboratory, Harvard School of Public Health, Boston, Massachusetts 02115, USA

§ Theodor Kocher Institute, University of Berne, Berne, CH-3012, Switzerland

Activated platelets bind numerous adhesive and procoagulant proteins by receptor-mediated processes¹. Although there is little evidence to suggest that these processes are heterogeneous in platelets, we previously found that platelets co-stimulated with collagen and thrombin express functional α -granule factor V only on a subpopulation of cells². Here we show that these cells, referred to as 'COAT-platelets', bind additional α -granule proteins, including fibrinogen, von Willebrand factor, thrombospondin, fibronectin and α_2 -antiplasmin. These proteins are all transglutaminase substrates, and inhibitors of transglutaminase prevent the production of COAT-platelets. A synthetic transglutaminase substrate (CP₁₅) also binds to COAT-platelets, and analysis by high performance liquid chromatography/mass spectrometry shows that a product is formed with a relative molecular mass (*M_r*) equal to CP₁₅ plus 176. Serotonin, an abundant component of platelet-dense granules³, has an *M_r* of 176, and fibrinogen isolated from COAT-platelets contains covalently linked serotonin. Synthetic bovine serum albumin-(serotonin)₆ binds selectively to COAT-platelets and also inhibits the retention of procoagulant

proteins on COAT-platelets. These data indicate that COAT-platelets use serotonin conjugation to augment the retention of procoagulant proteins on their cell surface through an as yet unidentified serotonin receptor.

Transglutaminases catalyse the formation of iso-peptide bonds between specific glutamine residues (acyl-donor) and either lysine or simple amines (amino-donor)^{4,5}. Platelets contain both tissue transglutaminase⁶ and factor XIII (ref. 7). Platelet serotonin in dense granules is released on cell activation and has two known roles: vasoconstriction and platelet activation¹. Convulxin is a snake venom protein that activates platelets through glycoprotein VI, a collagen receptor⁸.

We found that gel-filtered, human platelets activated with thrombin plus collagen, or thrombin plus convulxin, expressed high levels of surface-bound, α -granule factor V (FV) on about 30% of cells² (COAT-platelets; Fig. 1a, region M1) despite the fact that all platelets were activated². Further examination of COAT-platelets showed additional surface-bound, α -granule proteins, including fibrinogen/fibrin, von Willebrand factor, fibronectin, α_2 -antiplasmin and thrombospondin, but not IgG or albumin (Fig. 1b–h). The absolute numbers of FV, fibrinogen and thrombospondin on COAT-FV platelets were $4,200 \pm 300$; $21,100 \pm 2,900$; and $7,100 \pm 1,200$ (mean $\pm$ s.d.; $n = 4$) molecules per platelet, respectively (Supplementary Information).

Dual stimulation of platelets with ionophore A23187 plus thrombin generated more than 95% COAT-platelets (Fig. 1i), indicating that all platelets can become COAT-platelets. Therefore, the partial generation of COAT-platelets (Fig. 1a–f) does not represent the absence of an essential component in the non-COAT-platelet population; however, the reason for two platelet populations after activation with thrombin plus convulxin remains unclear.

As the α -granule proteins on the surface of COAT-platelets

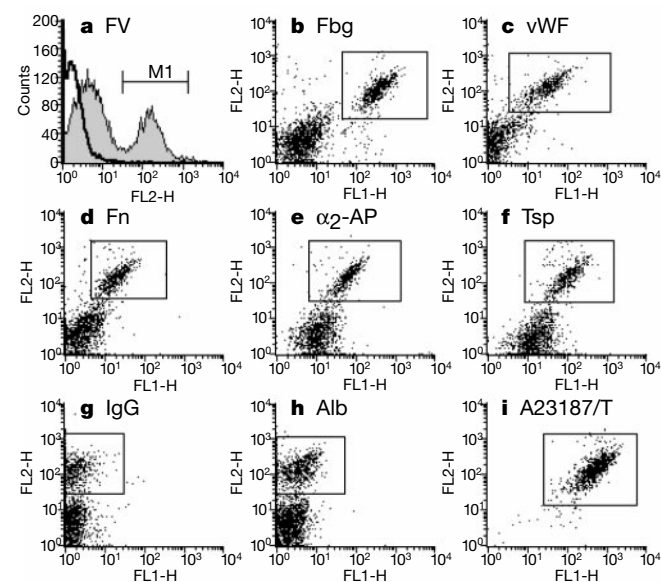


Figure 1 α -Granule proteins are concentrated on the surface of COAT-platelets. Gel-filtered platelets were activated with convulxin plus thrombin and stained for α -granule proteins on the cell surface. **a**, Binding of anti-FV antibody to control (line histogram) and activated (filled histogram) cells shows COAT-platelets (region M1) with a high concentration of bound FV (ref. 2). **b–h**, Analyses both for FV (FL2, ordinate) and additional α -granule proteins (FL1, abscissa): fibrinogen (**b**), von Willebrand factor (**c**), fibronectin (**d**), α_2 -antiplasmin (**e**), thrombospondin (**f**), IgG (**g**) and albumin (**h**). **i**, Platelets activated with ionophore A23187 plus thrombin and stained for FV (FL2) and fibrinogen (FL1). Boxes in **b–i** represent the COAT-platelets.

(Fig. 1) were all potential acyl-donors for transglutaminases^{4,9}, we tested whether the generation of COAT-platelets was affected by transglutaminase inhibitors. Dansylcadaverine, a competitive amino-donor for transglutaminases¹⁰, inhibited the appearance of FV, von Willebrand factor, fibrinogen and α_2 -antiplasmin on COAT-platelets (Fig. 2a), even though it had no apparent effect on platelet activation as measured by α -granule release (data not shown). The apparent half-maximal inhibitory concentration (IC_{50}) for dansylcadaverine was similar for each of these proteins, suggesting that it might be competing against a common amino-donor. In contrast, acetyl-casein, a competitive acyl-donor, showed distinct IC_{50} values for each of three acyl-donor substrates (Fig. 2b), inferring that these acyl-donor substrates compete against each other for incorporation into COAT-platelets. The generation of COAT-platelets was also prevented by SU732, an active site inhibitor of transglutaminases¹¹, with an IC_{50} of 160 μ M.

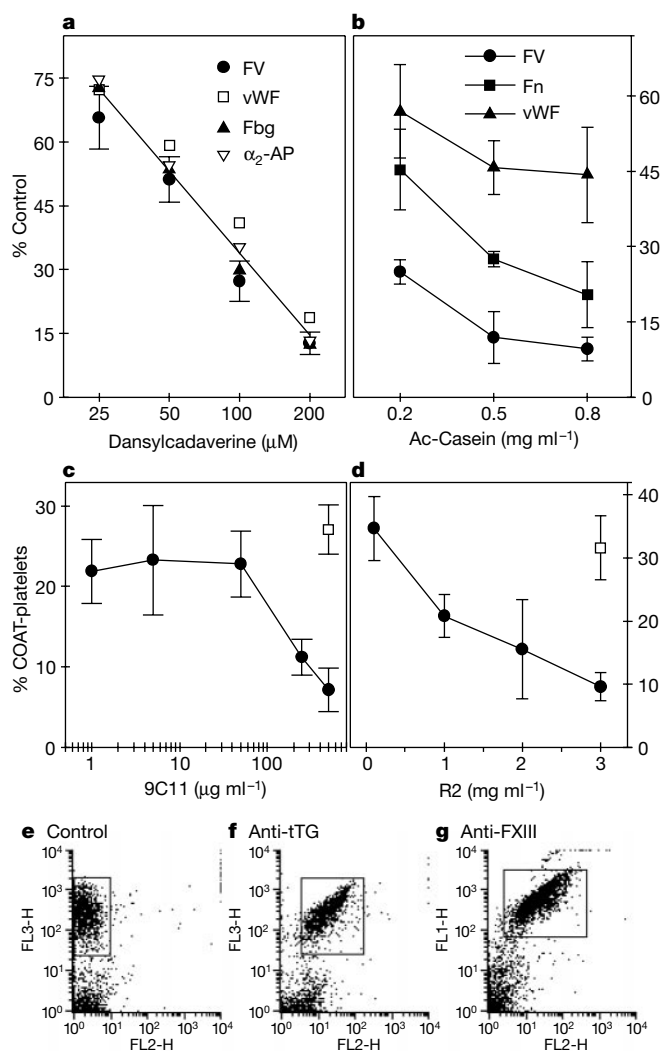


Figure 2 Dansylcadaverine, acetyl-casein and anti-FXIII inhibit COAT-platelet production. Platelets were activated with convulxin and thrombin in the presence of various inhibitors and then stained for α -granule proteins. **a**, Dansylcadaverine (μ M) inhibits equally the retention of FV, von Willebrand factor, fibrinogen or α_2 -antiplasmin (mean; $n = 3–5$) on COAT-platelets. **b**, Acetyl-casein ($mg\ ml^{-1}$) inhibits the incorporation of three different proteins into COAT-platelets (mean $\pm$ s.d., $n = 3$). **c**, **d**, Effect of anti-FXIII antibodies 9C11 and R2, respectively, on COAT-platelet synthesis as measured by FV incorporation (mean $\pm$ s.d., $n = 3–7$); open symbols represent control antibodies. **e–g**, Staining of dual-activated platelets with isotype control antibody (**e**), anti-tissue transglutaminase (**f**) and anti-FXIII (**g**); ordinates represent FV binding.

We tested two antibodies that inhibit factor XIII for their effect on COAT-platelets. Monoclonal antibody 9C11 inhibits thrombin activation of factor XIII (ref. 12), and R2 inhibits factor XIIIa activity¹³. These antibodies did not affect α -granule secretion or phosphatidylserine exposure (data not shown); however, both antibodies prevented, dose-dependently, the appearance of FV on COAT-platelets (Fig. 2c, d). Accumulation of fibrinogen and von Willebrand factor on these cells was also inhibited by 9C11 and R2 (data not shown). Although the high concentrations of antibody required for inhibition might raise doubts about their specificity, similar levels were required for inhibiting factor XIII in whole plasma¹². Unexpectedly, experiments with antibodies against tissue transglutaminase and factor XIII indicated that both transglutaminases are present on the surface of COAT-platelets (Fig. 2e–g). Further experiments will be needed to delineate the roles of these transglutaminases in COAT-platelet formation.

We synthesized a synthetic peptide (CP₁₅; SVLSLSQSKVLPVPE) that incorporates the transglutaminase-reactive glutamine (acyl-donor) of β -casein¹⁴. The biotinylated peptide bound to COAT-platelets (Fig. 3a, inset) in a manner similar to that observed for α -granule proteins, and this reaction was inhibited by dansylcadaverine (data not shown). Not only does this result show the direct coupling of a transglutaminase substrate to COAT-platelets, but it also offers an experimental approach to identify the platelet component that serves as the amino-donor.

We analysed biotin-CP₁₅ bound on COAT-platelets by SDS–

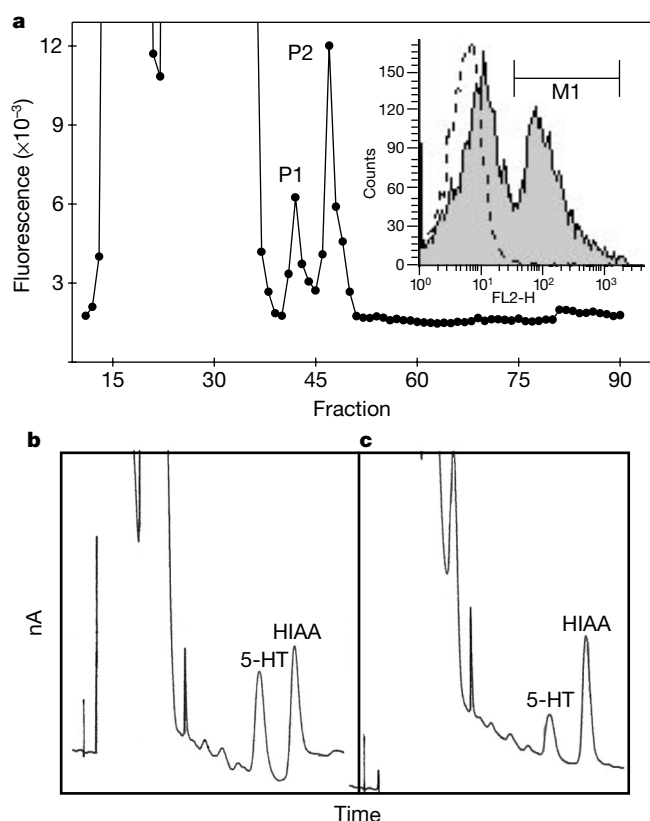


Figure 3 Serotonin conjugation to CP₁₅ and fibrinogen. Platelets were activated with convulxin plus thrombin in the presence of 100 μ M biotin-CP₁₅ or fl-CP₁₅. **a**, HPLC analysis of fl-CP₁₅ extracted from COAT-platelets (see Methods). Two fluorescent peaks not present in the control chromatography are indicated as P1 and P2. Inset, biotin-CP₁₅ binding to activated (filled histogram) and quiescent platelets (dashed line). M1 indicates COAT-platelets. **b**, **c**, HPLC analyses of hydrolysed fibrinogen isolated, respectively, from COAT- and quiescent-platelets. 5-Hydroxyindoleacetic acid (HIAA) was used as an internal control for hydrolysis and analysis procedures. nA, nanoamperes.

polyacrylamide gel electrophoresis (PAGE) and western blots, but observed no biotin-CP₁₅–protein complexes. Next, COAT-platelets were formed in the presence of 5(6)-carboxy-fluorescein-CP₁₅ (fl-CP₁₅), extracted with methanol, and analysed on a C₄ reverse phase high performance liquid chromatography (HPLC) column with gradient elution. We observed two peaks (P1 and P2) that were not present in the starting material (Fig. 3a), and analysed these by mass spectrometry. The predominant product, peak P2, represents the CP₁₅ substrate plus an M_r of 160. Allowing for the transglutaminase reaction, the M_r of the amino-donor was calculated to be 176. Serotonin (5-hydroxytryptamine; 5-HT), an abundant platelet constituent with an M_r of 176, is a recognized transglutaminase substrate¹⁵. Peak P1 represents a methylamine derivative of fl-CP₁₅ (P.F. and G.L.D., unpublished data).

Fibrinogen was immunopurified from control and COAT-platelets and hydrolysed with mercaptoethane sulphonic acid to release conjugated serotonin¹⁶. We then used reverse phase HPLC, in combination with electrochemical detection, to quantify the amount of serotonin. Fibrinogen from COAT-platelets contained conjugated serotonin, with an average of 7.6 ± 3.4 serotonin molecules per fibrinogen (mean $\pm$ s.d.; $n = 4$; Fig. 3b). Serotonin was only observed after acid hydrolysis, indicating that it does not represent non-covalently bound material. Unexpectedly, fibrinogen from control platelets also had 4.0 ± 1.8 serotonin molecules per fibrinogen (Fig. 3c), but this is significantly less than observed after activation ($P < 0.04$). Fibrinogen purified from plasma did not have detectable amounts of bound serotonin (data not shown).

To confirm the conjugation of serotonin to procoagulant proteins, we fractionated total soluble proteins from control and COAT-platelets using an HPLC molecular sieve column, and

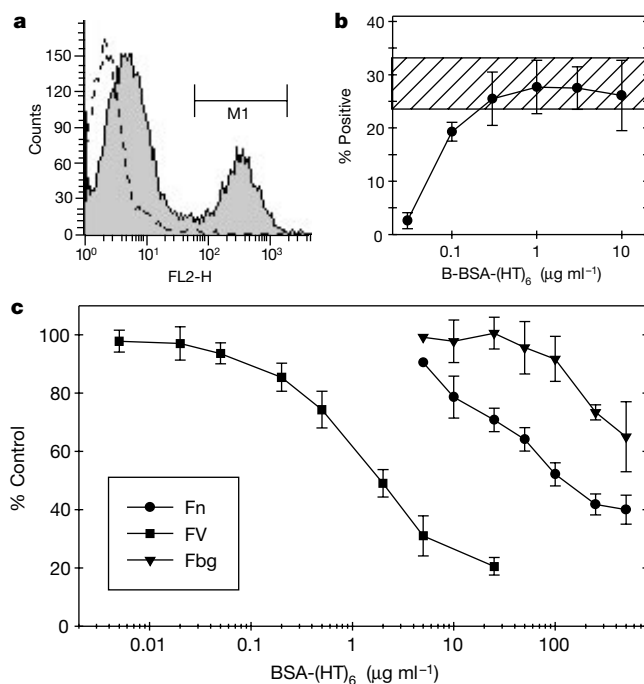


Figure 4 Binding of biotin-BSA-(5-HT)₆ to activated platelets. Platelets were activated with convulxin plus thrombin in the presence of 5 μ g ml⁻¹ biotin-BSA-(5-HT)₆ (B-BSA-(HT)₆); cells were analysed for bound biotin. **a**, Resting (dashed line) and activated (filled histogram) platelets. **b**, Percentage of platelets binding biotin-BSA-(5-HT)₆ (mean $\pm$ s.d., $n = 3$); hatched area represents COAT-platelet concentration (mean $\pm$ s.d.) for donors used in these experiments. **c**, Inclusion of BSA-(5-HT)₆ in incubations measuring the incorporation of proteins into COAT-platelets. The decrease in fluorescence (ordinate) with graded doses of BSA-(5-HT)₆ is plotted for FV, fibrinogen and fibronectin on COAT-platelets (see also Supplementary Information).

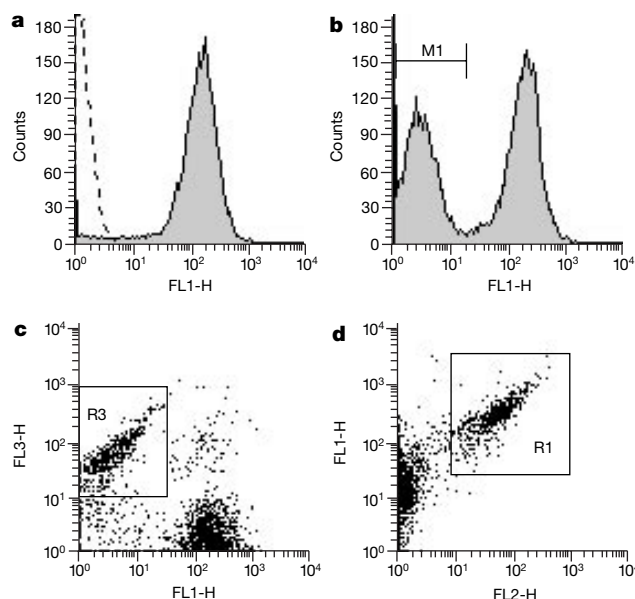


Figure 5 COAT-platelets bind low levels of PAC-1. **a, b**, Platelets activated with thrombin alone (**a**; filled histogram) or convulxin plus thrombin (**b**) in the presence of FITC-PAC-1. **c**, Analysis of both FV (ordinate) and FITC-PAC-1 on convulxin plus thrombin activated

platelets shows that COAT-platelets (R3) bind low levels of PAC-1. **d**, Cells activated by convulxin plus thrombin and stained with LIBS-6 (FL1) and anti-fibronectin (FL2); cells in R1 are positive for both probes.

analysed the high molecular weight fractions for conjugated serotonin as above. Both control and COAT-platelet extracts had serotonin conjugated to the macromolecular fraction. COAT-platelets averaged 1.8 ± 0.3 times ($n = 3$) more serotonin, roughly the same fold increase observed with purified fibrinogen. Fibrinogen-serotonin conjugates in resting platelets may either represent derivatization occurring as fibrinogen is packaged into α -granules or, alternatively, selective uptake of very low amounts of fibrinogen-serotonin adducts from the plasma.

We synthesized a serotonin adduct of CP₁₅, biotin-CP₁₅-Q⁷-5'-HT (B-CP₁₅-(5-HT)), using tissue transglutaminase. Platelets activated with convulxin plus thrombin bound B-CP₁₅-(5-HT), whereas resting cells did not (Supplementary Information). A multivalent serotonin adduct of bovine serum albumin, biotin-BSA-(5-HT)₆, was synthesized chemically and tested for its ability to bind to COAT-platelets. Biotin-BSA-(5-HT)₆ bound to dual-agonist-activated platelets in a concentration-dependent manner, and maximal binding closely approximated the percentage of COAT-platelets (Fig. 4a, b). Furthermore, BSA-(5-HT)₆ inhibited the retention of α -granule proteins on COAT-platelets. FV, fibrinogen and fibronectin concentrations on COAT-platelets were dose-dependently attenuated by BSA-(5-HT)₆ (Fig. 4c); the IC₅₀ for inhibition of FV was $1.8 \mu\text{g ml}^{-1}$ or roughly 26 nM. Furthermore, a serotonin transport inhibitor, fluoxetine¹⁷, inhibited the retention of FV and fibrinogen by COAT-platelets (Supplementary Information).

As the α -granule proteins retained on COAT-platelets seemed to be anchored through transglutaminase-mediated addition of serotonin, we considered whether these proteins were also binding to their respective receptors. We used PAC-1, a monoclonal antibody that recognizes the fibrinogen-binding site of glycoprotein IIb/IIIa¹⁸, to determine whether glycoprotein IIb/IIIa receptors on COAT-platelets were free or occupied. All platelets activated with thrombin (or convulxin) alone bound PAC-1 (Fig. 5a); however, activation with thrombin plus convulxin in the presence of PAC-1 produced two platelet populations (Fig. 5b). Staining for surface glycoprotein IIb/IIIa after thrombin plus convulxin activation showed that all platelets had a similar receptor density (data not shown). Labelling for FV and PAC-1 after dual-agonist activation (Fig. 5c) showed that COAT-platelets (region R3) are the cells that bind low levels of fluorescein isothiocyanate (FITC)-PAC-1. We

also stained the dual-activated platelets with LIBS-6, a monoclonal antibody that recognizes glycoprotein IIb/IIIa with bound ligand¹⁹. COAT-platelets, identified by bound fibronectin, also react with LIBS-6 (Fig. 5d). These results show that the glycoprotein IIb/IIIa receptors on COAT-platelets are occupied by ligand.

It is noteworthy that PAC-1 cannot displace ligands from COAT-platelets. Considering that PAC-1 has a 50-fold higher affinity for activated glycoprotein IIb/IIIa than has fibrinogen¹⁸, and that bound fibrinogen and fibronectin on COAT-platelets are also not displaced by DMP-802 (G.L.D., unpublished data), another strong glycoprotein IIb/IIIa antagonist²⁰, it is reasonable to suggest that glycoprotein IIb/IIIa receptors on COAT-platelets are retaining adhesive proteins by an unknown mechanism.

These observations are compatible with a model in which adhesive and procoagulant proteins in α -granules are conjugated with serotonin on platelet activation, and serotonin-protein adducts then bind to an as yet unidentified serotonin receptor on the cell surface. This interaction augments and stabilizes the normal receptor-mediated binding of adhesive and procoagulant proteins to the platelet, for example, fibrinogen to glycoprotein IIb/IIIa or FV to phosphatidylserine. As a result, COAT-platelets represent a unique component of haemostasis. The requirement for dual stimulation indicates that COAT-platelets will only be formed under circumstances of extreme haemostatic need, such as immobilization of platelets on the collagen surface of a ruptured vessel in the presence of continuous thrombin generation. Moreover, COAT-platelets possessing a high surface density of FV and adhesive proteins seem well suited for generating prothrombinase complexes as well as serving as focal points for platelet-platelet adhesion. Alternatively, COAT-platelets may represent a thrombotic risk if elicited under improper circumstances. Clarification of the exact physiological role of COAT-platelets awaits further investigation. □

Methods

Materials

Sephacore CL-2B, FITC-goat-anti-mouse-IgG (FITC-GAMG), bovine thrombin, 5-HT, 5-hydroxyindoleacetic acid, A23187, goat-anti-human fibrinogen, monoclonal-anti-fibrinogen (85D4), rabbit-anti-human fibronectin, goat-anti-human-IgG, goat-anti-human-albumin, guinea-pig liver transglutaminase, bovine casein, BSA, MES and HEPES

buffers were obtained from Sigma; phycoerythrin-labelled streptavidin (PE-SA) from Molecular Probes; mercaptoethane sulphonic acid from Pierce; and FITC-PAC-1 from Becton-Dickinson.

We used the following monoclonal antibodies: G5 (anti-P-selectin) and TAB (anti-glycoprotein IIb/IIIa) from R. P. McEver; HFV237 and HFV271 against FV from C. T. Esmon; LIBS-6 (ref. 19) from M. Ginsberg; ED4H1 against human α_2 -antiplasmin (α_2 -AP) from P. A. McKee; monoclonal 9C11, inhibiting thrombin activation of factor XIII (ref. 12); and CUB7402 against tissue transglutaminase and C6.7 against thrombospondin from Neomarkers. R2 is a purified, polyclonal antibody raised against plasma factor XIII that inhibits the transglutaminase activity of factor XIIIa¹³; goat-anti-human-factor XIII for flow cytometry was from Biodesign International; rabbit anti-human-thrombospondin was from Calbiochem. SU732 (1,3,4,5-tetramethyl-2-[(2-oxopropyl)thio]imidazolium chloride; compound I)¹¹ and DMP-802 (ref. 20) were from A. Stern. Convulxin was purified as described⁸. A 15-amino-acid peptide (SVLSLSQSKVLPVPE; CP₁₅) representing residues 161–175 of β -casein¹⁴ and thrombin receptor agonist peptide (TRAP; SFLLRN) were synthesized by the Molecular Biology Resource Facility of the Univ. Oklahoma Health Sciences Center. Acetylated casein (acetyl-casein)²¹ and buffers² were prepared as described.

Platelet activation

Blood was drawn into acid citrate dextrose (ACD), and gel-filtered platelets were prepared and diluted to $4 \times 10^7 \text{ ml}^{-1}$ in buffered saline-glucose-citrate (BSGC) as described². Platelets were activated in 100- μl reactions containing 40 μl of HEPES/saline with 5 mM CaCl_2 and 2.5 mM MgCl_2 ; 40 μl of HEPES/saline with 1 mg ml^{-1} BSA, 10 μl of agonist, inhibitor and/or antibody, and 10 μl of gel-filtered platelets; standard convulxin and thrombin concentrations were 500 ng ml^{-1} and 5 nM final, respectively². After 10 min of activation at 37°C, 200 μl of 1.5% (v/v) formalin in PBS was added to stop the reaction and fix the platelets. After 20 min of fixation at room temperature, 4 ml of 1 mg ml^{-1} BSA in PBS were added, and the samples were centrifuged at 1,500g for 15 min. We resuspended platelets in 200 μl of BSA/PBS and stained them with the appropriate secondary reagents. Several antibodies were included during platelet activation, including anti-FV (HFV237), anti- α_2 -AP (ED4H1), LIBS-6 and anti-tissue transglutaminase (CUB-7402). Other antibodies were incubated with platelets after activation and fixation. For inhibition studies, we pre-incubated platelets with inhibitory antibodies against factor XIII (9C11 or R2) for 15 min at room temperature before platelet activation.

Quantification of bound proteins

Platelets were activated with convulxin plus thrombin in the presence of 10 $\mu\text{g ml}^{-1}$ of individual monoclonal antibodies: HFV271 (FV), TAB (glycoprotein IIb/IIIa), AP2 (glycoprotein IIb/IIIa), 85D4 (fibrinogen) and C6.7 (thrombospondin). After formalin fixation and washing, platelets were stained with 15 $\mu\text{g ml}^{-1}$ PE-goat-anti-mouse IgG, washed and analysed at a constant voltage. Geometric mean fluorescence for each antibody was determined and referenced to AP2 binding to resting platelets²².

CP₁₅ adduct analysis

Gel-filtered platelets (2×10^9) were activated with A23187 plus thrombin² in the presence of 50 μM fl-CP₁₅; five volumes of methanol were added to stop the reaction, and the sample was centrifuged at 2000g for 15 min. We diluted the supernatant with water to achieve 10% (v/v, final) methanol, and applied it to a 10 $\times$ 25 mm C₄ column (Vydac) at a flow rate of 5 ml min^{-1} . The column was eluted at 1 ml min^{-1} with a 10–50% gradient of acetonitrile in 10 mM triethylamine/acetic acid (pH $\approx$ 7). We collected 1-ml fractions and analysed them for fluorescence (485-nm excitation/530-nm emission) with a Biotek FL600 fluorometer. Two product peaks (P1 and P2) were identified and the relevant fractions were applied to a SPE/C₄ desalting column (J. T. Baker) and eluted with 50% acetonitrile in water. The fractions were then lyophilized.

Electrospray mass spectrometry was performed with a Sciex QSTAR hybrid quadrupole time-of-flight mass spectrometer (Applied Biosystems) operated in the positive ion mode. We loaded 2 μl of sample dissolved in 0.5% acetic acid in methanol/water (1:1) into a nanospray capillary, which was then mounted in the nanospray source (MDS Protana). The capillary was opened and positioned 2 mm from the orifice of the mass spectrometer. A potential of 1,375 V was applied to the needle to initiate the nanospray process. Data were collected over 400–3,000 m/z units.

Synthesis of serotonin adducts

Biotin-CP₁₅ and fl-CP₁₅ (400 μM) were incubated with 10 mM serotonin, 5 mM CaCl_2 , 0.1 mM dithiothreitol, 200 mM HEPES (pH 7.5) and 0.2 mU ml^{-1} guinea-pig liver tissue transglutaminase for 30 min at 37°C. This reaction mixture was extracted with methanol and purified over HPLC as outlined above; fl-CP₁₅-(5-HT) co-migrated with peak P2 of Fig. 3a. We synthesized BSA-(5-HT)_n by reacting BSA (5 mg ml^{-1}) in 25 mM serotonin, 200 mM MES (pH 5.8) with 20 mM ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDAC) for 3 h at 37°C. We dialysed the product against HEPES/saline (pH 7.5), and measured the absorbance at 275 nm (A_{275}) to calculate the level of 5-HT substitution. BSA-(5-hydroxyindoleacetic acid)_n was synthesized similarly.

Analysis of fibrinogen-serotonin adducts

We activated 4×10^8 platelets with 1 μM A23187 and 25 μM TRAP for 10 min at 37°C. The reaction was stopped with 5 mM EDTA, 0.1 mM phenylmethylsulphonyl fluoride and 1% Triton X-100 (final). Fibrinogen from activated and control samples was bound to anti-fibrinogen-Sepharose beads and subsequently eluted with 8 M urea by standard

techniques. Recovered fibrinogen was dialysed against 10 mM NaCl, 0.1 mM EDTA, quantified with an enzyme-linked immunoabsorbent assay, and lyophilized. We spiked samples with a known concentration of BSA-(5-hydroxyindoleacetic acid) to serve as an internal control for hydrolysis and recovery of serotonin adducts. Generally, 0.2–1 μg of fibrinogen was analysed, and 2–20 pmol of serotonin was detected. Samples were hydrolysed with 3 M mercaptoethane sulphonic acid containing 2 mg ml^{-1} tryptophan for 20 min at 160°C (ref. 16). Hydrolysed samples were analysed on a Zorbax 4.6 $\times$ 250 mm, 5- μm C₁₈ column (Agilent Technologies), equilibrated with 5 mM hexanesulphonic acid, 0.1 mM EDTA, 100 mM sodium phosphate (pH 3) with 7% (v/v) acetonitrile²³; a Coulchem II electrochemical detector (ESA) was used to measure serotonin and 5-hydroxyindole acetic acid in the column eluate.

Received 3 September; accepted 30 October 2001.

- Ware, J. A. & Collier, B. S. In *Williams Hematology* (eds Beutler, E., Lichtman, M. A., Collier, B. S. & Kipps, T. J.) 1161–1201 (McGraw-Hill, New York, 1995).
- Alberio, L., Safa, O., Clemetson, K. J., Esmon, C. T. & Dale, G. L. Surface expression and functional characterization of α -granule factor V in human platelets: Effects of ionophore A23187, thrombin, collagen and convulxin. *Blood* **95**, 1694–1702 (2000).
- McNicol, A. & Israels, S. J. Platelet dense granules: structure, function and implications for haemostasis. *Thromb. Res.* **95**, 1–18 (1999).
- Greenberg, C. S., Birckbichler, P. J. & Rice, R. H. Transglutaminases: multifunctional cross-linking enzymes that stabilize tissues. *FASEB J.* **5**, 3071–3077 (1991).
- Pincus, J. H. & Walsch, H. The specificity of transglutaminase. II. Structural requirements of the amine substrate. *Arch. Biochem. Biophys.* **126**, 44–52 (1968).
- Puszkin, E. G. & Raghuraman, V. Catalytic properties of a calmodulin-regulated transglutaminase from human platelet and chicken gizzard. *J. Biol. Chem.* **260**, 16012–16020 (1985).
- Muszbec, L., Adany, R. & Mikkola, H. Novel aspects of blood coagulation factor XIII. I. Structure, distribution, activation, and function. *Crit. Rev. Clin. Lab. Sci.* **33**, 357–421 (1996).
- Polgar, J. et al. Platelet activation and signal transduction by convulxin, a C-type lectin from *Crotalus durissus terrificus* (tropical rattlesnake) venom via the p62/GPVI collagen receptor. *J. Biol. Chem.* **272**, 13576–13583 (1997).
- Aeschlimann, D. & Paulsson, M. Transglutaminases: Protein cross linking enzymes in tissues and body fluids. *Thromb. Haemost.* **71**, 402–415 (1994).
- Parameswaran, K. N., Velasco, P. T., Wilson, J. & Lorand, L. Labeling of ϵ -lysine crosslinking sites in proteins with peptide substrates of factor XIIIa and transglutaminase. *Proc. Natl Acad. Sci. USA* **87**, 8472–8475 (1990).
- Lorand, L., Stern, A. M. & Velasco, P. T. Novel inhibitors against the transglutaminase-catalysed crosslinking of lens proteins. *Exp. Eye Res.* **66**, 531–536 (1998).
- Reed, G. L. & Lukacova, D. Generation and mechanism of action of a potent inhibitor of factor XIII function. *Thromb. Haemost.* **74**, 680–685 (1995).
- Reed, G. L., Matsueda, G. R. & Haber, E. Platelet factor XIII increases the fibrinolytic resistance of platelet rich clots by accelerating the crosslinking of α -antiplasmin to fibrin. *Thromb. Haemost.* **68**, 315–320 (1992).
- Gorman, J. J. & Folk, J. E. Structural features of glutamine substrates for transglutaminases. *J. Biol. Chem.* **256**, 2712–2715 (1981).
- Clarke, D. D., Mycek, M. J., Neidle, A. & Walsch, H. The incorporation of amines into protein. *Arch. Biochem. Biophys.* **79**, 338–354 (1959).
- Maeda, K., Scheffler, J. J. & Tsugita, A. Tryptophan micro-scale determinations by rapid hydrolysis. *Hoppe-Seyler's Z. Physiol. Chem.* **365**, 1183–1185 (1984).
- Wagner, A., Montero, D., Martensson, B., Siwers, B. & Asberg, M. Effects of fluoxetine treatment of platelet 3H-imipramine binding, 5-HT uptake and 5-HT content in major depressive disorder. *J. Affect. Disord.* **20**, 101–113 (1990).
- Shattil, S. J., Cunningham, M. & Hoxie, J. A. Detection of activated platelets in whole blood using activation-dependent monoclonal antibodies. *Blood* **70**, 307–315 (1987).
- Huang, M. M. et al. Adhesive ligand binding to integrin αIIb β 3 stimulates tyrosine phosphorylation of novel protein substrates before phosphorylation of pp125FAK. *J. Cell Biol.* **122**, 473–483 (1993).
- Mousa, S. A. et al. Oral antiplatelet efficacy and specificity of a novel nonpeptide platelet GPIIb/IIIa receptor antagonist, DMP 802. *J. Cardiovasc. Pharmacol.* **32**, 169–176 (1998).
- Means, G. E. & Feeney, R. E. *Chemical Modification of Proteins* (Holden-Day, San Francisco, 1971).
- Pidard, D., Montgomery, R. R., Bennett, J. S. & Kunicki, T. J. Interaction of AP-2, a monoclonal antibody specific for the human platelet glycoprotein IIb-IIIa complex, with intact platelets. *J. Biol. Chem.* **258**, 12582–12586 (1983).
- Laakso, J. T., Koskineemi, M. L., Wahlroos, O. & Harkonen, M. Simultaneous determination of tryptophan and its 5-hydroxy metabolites in human cerebrospinal fluid by reversed phase liquid chromatography with electrochemical detection. *Scand. J. Clin. Lab. Invest.* **43**, 463–472 (1983).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

This work was supported in part by an NIH grant (G.L.D.), the W.K. Warren Medical Research Institute (G.L.D.), Novartis (L.A.) and the Swiss National Science Foundation (K.J.C.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.L.D. (e-mail: george-dale@ouhsc.edu).

Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry

Yuen Ho*, Albrecht Gruhler*, Adrian Heilbut*, Gary D. Bader†‡, Lynda Moore*, Sally-Lin Adams*, Anna Millar*, Paul Taylor*, Keiryn Bennett*, Kelly Boutilier*, Lingyun Yang*, Cheryl Wolting*, Ian Donaldson*, Søren Schandorff*, Juanita Shewnarane*, Mai Vo†‡, Joanne Taggart†‡, Marilyn Goudreau†‡, Brenda Muskat*, Cris Alfaro*, Danielle Dewar†, Zhen Lin†, Katerina Michalickova†‡, Andrew R. Willems†§, Holly Sassi†, Peter A. Nielsen*, Karina J. Rasmussen*, Jens R. Andersen*, Lene E. Johansen*, Lykke H. Hansen*, Hans Jespersen*, Alexandre Podtelejnikov*, Eva Nielsen*, Janne Crawford*, Vibeke Poulsen*, Birgitte D. Sørensen*, Jesper Matthiesen*, Ronald C. Hendrickson*, Frank Gleeson*, Tony Pawson†§, Michael F. Moran*, Daniel Durocher†§, Matthias Mann*, Christopher W. V. Hogue†‡, Daniel Figey* & Mike Tyers†§

* MDS Proteomics, 251 Attwell Drive, Toronto, Canada M9W 7H4, and Staermosegaardsvej 6, DK-5230 Odense M, Denmark

† Programme in Molecular Biology and Cancer, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Canada, M5G 1X5

‡ Department of Biochemistry, University of Toronto, 1 Kings College Circle, Toronto, Canada M5S 1A8

§ Department of Medical Genetics and Microbiology, University of Toronto, 1 Kings College Circle, Toronto, Canada M5S 1A8

The recent abundance of genome sequence data has brought an urgent need for systematic proteomics to decipher the encoded protein networks that dictate cellular function¹. To date, generation of large-scale protein–protein interaction maps has relied on the yeast two-hybrid system, which detects binary interactions through activation of reporter gene expression^{2–4}. With the advent of ultrasensitive mass spectrometric protein identification methods, it is feasible to identify directly protein complexes on a proteome-wide scale^{5,6}. Here we report, using the budding yeast *Saccharomyces cerevisiae* as a test case, an example of this approach, which we term high-throughput mass spectrometric protein complex identification (HMS-PCI). Beginning with 10% of predicted yeast proteins as baits, we detected 3,617 associated proteins covering 25% of the yeast proteome. Numerous protein complexes were identified, including many new interactions in various signalling pathways and in the DNA damage response. Comparison of the HMS-PCI data set with interactions reported in the literature revealed an average threefold higher success rate in detection of known complexes compared with large-scale two-hybrid studies^{3,4}. Given the high degree of connectivity observed in this study, even partial HMS-PCI coverage of complex proteomes, including that of humans, should allow comprehensive identification of cellular networks.

To survey the yeast proteome, we chose an initial set of 725 bait proteins representative of a variety of different functional classes, including 100 protein kinases, 36 phosphatases and regulatory subunits, and 86 proteins implicated in the DNA damage response (DDR). A small scale, one-step immuno-affinity purification based on the Flag epitope tag was used to capture bait proteins, which were transiently overexpressed from the heterologous *GAL1* or *tet* promoters. Proteins from 1,558 individual immunoprecipitations were resolved by SDS–polyacrylamide gel electrophoresis (PAGE), visualized by colloidal Coomassie stain, excised from the gel and subjected to tryptic digestion before mass spectrometric analysis (Supplementary Information Fig. 1). As our isolation procedure often yielded multiple proteins from single excised bands, which

cannot be resolved by peptide-mass-fingerprinting alone, we used tandem mass spectrometry (MS/MS) fragmentation to identify unambiguously proteins in each gel slice⁶. A total of 15,683 gel slices were processed, yielding approximately 940,000 MS/MS spectra that matched sequences in the protein sequence database. Over 35,000 protein identifications were made, corresponding to 8,118 potential interactions with a set of 600 bait proteins that were expressed at detectable levels (Supplementary Information Table 1). Ubiquitous, nonspecifically binding proteins, defined empirically on the basis of frequency of occurrence (see Supplementary Information), were subtracted from the raw data set to yield 3,617 interactions with 493 baits—for further discussion of filtering criteria and mass spectrometry methodology, see Supplementary Information. This filtered data set contained 1,578 different interacting proteins representing 25% of the yeast proteome (Supplementary Information Table 2). In a preliminary direct validation of the HMS-PCI data set, 64 out of 86 interactions (74%) in a random set of new associations detected by HMS-PCI were recapitulated in immunoprecipitation-immunoblot experiments (data not shown). The HMS-PCI method was able to identify known complexes from a variety of subcellular compartments, including the cytoplasm, cytoskeleton, nucleus, nucleolus, plasma membrane, mitochondrion and vacuole (see Supplementary Information). Of all the proteins identified, 531 corresponded to hypothetical uncharacterized proteins predicted from the yeast genome sequence (Supplementary Information Table 3).

To begin to assess cellular signalling events on a proteome-wide level, we used most of the protein kinases and phosphatases encoded in the yeast genome to capture associated components. As an example, HMS-PCI analysis of the mitogen-activated protein kinase (MAPK) Kss1 identified many known components of the mating/filamentous growth pathway⁷, including Ste11, Ste7, and four known downstream targets, the transcriptional regulators, Ste12, Tec1, Dig1/Rst1 and Dig2/Rst2 (Fig. 1a). We identified other Kss1 interactions of potential biological significance, including Bem3, which is a GTPase-activating protein that may help attenuate the upstream Cdc42 Rho-type GTPase needed for mating. Relevant interactions were also detected between the cell wall integrity MAPK Slt2 and its upstream activators Bck1 and Mkk2, and between the osmotic stress response MAPK Hog1 and Rck2, a downstream target kinase⁷ (Supplementary Information Table 2).

Numerous kinases and phosphatases formed discernable complexes, often with regulatory factors that serve to localize or control activity¹. An extensive network was assembled around Cdc28, the primary cyclin-dependent kinase (CDK) for cell division control⁸, including interactions with its known cyclin partners Cln1, Cln2, Clb2, Clb3 and Clb5, as well as with the small CDK-binding subunit Cks1 (Fig. 1b). In many cases the detected interactions are bridged by intermediary partners; for example crystallographic evidence indicates that Cks1–cyclin interactions occur through the CDK

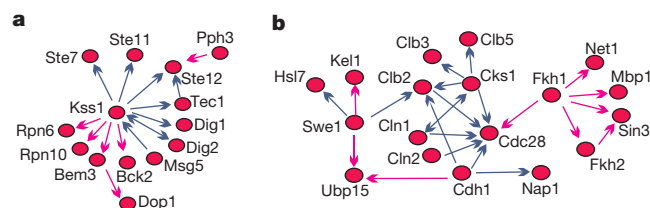


Figure 1 Two kinase-based signalling networks. **a**, Interaction diagram for Kss1 complexes. **b**, Interaction diagram for Cdc28 and Fkh1/2 complexes. Arrows point from the bait protein to the interaction partner. Blue arrows indicate known interactions; red arrows indicate new interactions. Not all detected interactions with all components are shown, nor are connections necessarily direct (see Supplementary Information Table 2).

subunit⁸. However, in the absence of additional evidence, there is no a priori means to elucidate connectivity of the interactions, and so we have chosen to represent each interaction as a direct pair (see below). The dual-specificity kinase Swe1, which inhibits Cdc28, was associated with both Clb2 and Hsl7, a known negative regulator of Swe1 (ref. 9). An interaction between Swe1 and Kell1, a protein that is involved in cell fusion and cell polarity¹⁰, might reflect the link between polarized growth and Swe1 activity. Cdc28 also associated with Fkh1, one of the transcription factors that drives expression of the mitotic cyclins *CLB1/2* and other G2/M-regulated genes, thereby providing direct physical closure of the known Clb1/2-Cdc28 positive feedback loop¹¹. Other Fkh1/2 interactions were consistent with roles in transcriptional activation and repression. Another local network was detected around the mitotic exit network (MEN) kinases Cdc5, Cdc15, Dbf2 and Dbf20, including the cohesin complex recently implicated as a Cdc5 target for activation of sister chromatid separation¹². Finally, numerous interactions between protein phosphatases, regulatory subunits and substrates were found (Supplementary Table S2). For example, the dual specificity phosphatase Msg5 was associated with both Fus3 and Slt2, consistent with its genetic role in attenuating MAPK signalling⁷.

The DDR includes DNA repair processes and checkpoint pathways that dictate cell cycle progression, transcription, protein degradation and DNA repair itself¹³. The global DDR network revealed by HMS-PCI contained many known interactions as well as many new interactions of probable biological significance (Fig. 2a; see also Supplementary Information Table 2). Most of the interactions identified did not depend on treatment with exogenous DNA-damaging agents, perhaps reflecting the fact that low level DNA damage normally occurs during replication¹³. Examples of known interactions include: the replication factor C complex (RFC, Rfc1–5) and the RFC^{Rad24} subcomplex, as well as the PCNA-like (PCNAL) Mec3–Rad17–Ddc1 complex, both of which transduce DNA damage signals; part of the Mms2–Ubc13–Rad18 post-replicative repair (PRR) complex; and the Mre11–Rad50–Xrs2 (MRX) complex that mediates double-strand-break repair by homologous and non-homologous mechanisms¹³. We also recovered nearly all known nucleotide excision repair (NER) factors in their dedicated subcomplexes¹⁴: Rad1–Rad10–Rad14 (NEF1); Rad3–TFB3–Kin28–Ccl1 (NEF3/TFIIH); and Rad7–Rad16 (NEF4). The Rad4–Rad23 interaction (NEF2) was not found, but we nevertheless detected an association between Rad4 and NEF1, a known interaction among NER factors.

The comprehensive coverage of DDR proteins readily identified pathway and network connections. For example, we recovered Rfc4 in Ddc1 complexes, consistent with the hypothesis that the PCNAL complex might be loaded onto DNA by the RFC^{Rad24} complex¹⁵. In terms of new interactions, the HMS-PCI approach revealed that Met18 can associate with Rad3, a component of the transcription factor TFIIF complex needed for both RNA Pol II-dependent transcription and NER. An association between Rad23 and the ubiquitin chain assembly factor Ufd2 (ref. 16) is corroborated by genetic interactions that suggest that *RAD23* and *UFD2* act antagonistically¹⁷. Ufd2 also interacted with a second ubiquitin-like (UBL) domain-containing protein, Dsk2. Another NER component, Rad7, bound the yeast elongin C homologue, Elc1, for which a function remains to be assigned. In mammalian cells, elongin C associates with elongin B, the cullin Cul2, the RING-H2 domain protein Rbx1 and substrate recruitment factors called SOCS box proteins to form E3 ubiquitin ligase complexes¹⁸. Consistent with the Elc1–Rad7 interaction, sequence alignments revealed a divergent SOCS box motif in Rad7, suggesting that it may mediate substrate ubiquitination during excision repair (Fig. 2b, c).

The Rad53 kinase, which corresponds to Chk2 in mammals and Cds1 in fission yeast, is a conserved mediator of DDR signals¹³. In

addition to the known Rad53 interaction with Asf1 (refs 19, 20), HMS-PCI revealed several new, associated proteins of probable biological significance, including the protein phosphatase 2C (PP2C) enzyme Ptc2, which is genetically implicated as a negative regulator of Rad53 (ref. 21). Furthermore, the uncharacterized gene

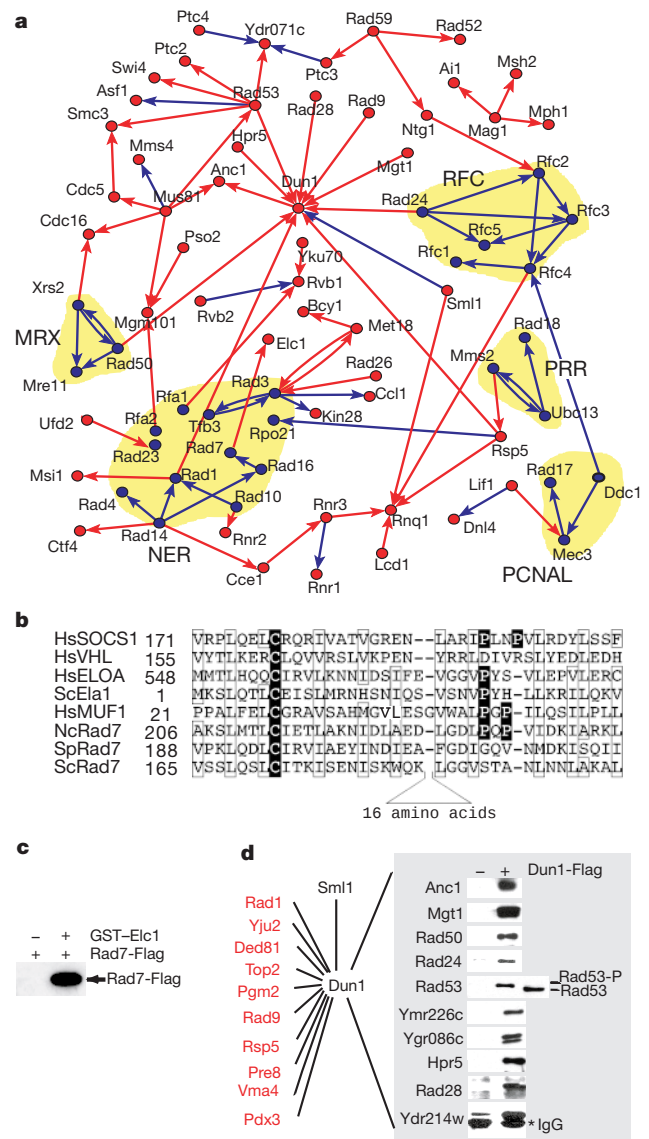


Figure 2 The DNA damage response network. **a**, Global connections in the network represented by a subset of known and new interactions. Interactions were initially nucleated from 86 proteins implicated in the DDR. Blue nodes indicate known interactions within dedicated complexes, highlighted in yellow. Blue arrows indicate known interactions; red arrows indicate new interactions. Not all detected interactions with all components are shown (see Supplementary Information Table 2), nor are connections necessarily direct. **b**, Alignment of SOCS box motif in Rad7 with known SOCS box proteins, including human MUF1, which like Rad7 contains carboxy-terminal leucine-rich repeats. Conserved hydrophobic residues are boxed; other diagnostic residues are shaded. Hs, *Homo sapiens*; Sc, *Saccharomyces cerevisiae*; Nc, *Neurospora crassa*; Sp, *Schizosaccharomyces pombe*. **c**, Rad7 interacts with Elc1. Flag-tagged Rad7 and glutathione *S*-transferase (GST)–Elc1 or GST alone were expressed in yeast, followed by capture on glutathione resin and detection of the Flag epitope. **d**, Complete set of Dun1 interactions detected by HMS-PCI. Co-immunoprecipitation tests of a subset of new Dun1 interactions are shown on the right; untested interactions are shown on the left. Note that the phosphorylated form of Rad53 interacts specifically with Dun1. The colour scheme is the same as in **a**. IgG, immunoglobulin-γ. Heavy chain; asterisk.

product Ydr071c was detected with both Rad53 and the PP2C family members Ptc3 and Ptc4, suggesting that Ydr071c may be a DDR-specific regulatory factor of PP2C-type phosphatases. We found consistently a genetic interaction between *YDR071C* and *RAD53* (R. Woolstencroft and D.D., unpublished data). With regard to Rad53 substrates¹³, the putative targets Swi4 and Cdc5 were linked directly or indirectly to Rad53 by HMS-PCI.

The Dun1 protein kinase has a similar overall structure to Rad53, most notably through the presence of phosphothreonine-binding modules termed forkhead-associated (FHA) domains²². Dun1 is implicated in many aspects of the DDR, yet the identities of its upstream regulators and downstream effectors remain largely unknown. HMS-PCI analysis of Dun1 revealed potential upstream regulators Rad9, Rad53, Rad24, Hpr5/Srs2 and Rad50 (Fig. 2d). Dun1 interacted with a probable substrate, the ribonucleotide reductase inhibitor Sml1, which is degraded on *DUN1*-dependent phosphorylation²³. Interestingly, Dun1 also associated with Rsp5, an E3 ubiquitin ligase that targets the RNA polymerase II large subunit Rpo21 for ubiquitin-mediated degradation after DNA damage²⁴. Rsp5 is thus a candidate E3 ubiquitin ligase for Sml1. Of the interactions tested by co-immunoprecipitation, 10 out of the 10 that involved Dun1 were confirmed. Of particular significance are the interactions between Dun1 with Rad24, Rad50, Rad53,

Hpr5/Srs2, Rad28 and Mgt1, as all of these proteins are known to participate in the DDR. We note that Dun1 was able to retrieve selectively the hyperphosphorylated form of Rad53, probably through the FHA domain of Dun1 (Fig. 2d). This observation suggests that the DNA damage signal is transmitted directly from Rad53 to Dun1. Finally, Dun1 also interacts with Ymr226c and Ygr086c, two proteins of unknown function that are induced on general cell stress, perhaps indicating that Dun1 may affect processes other than DNA damage.

We compared the HMS-PCI data set, represented as hypothetical direct interactions between bait and associated proteins, with comprehensive high-throughput yeast two-hybrid (HTP-Y2H) data sets^{3,4} using interactions reported in the literature as a benchmark. Interaction data sets were entered into the Biomolecular Interaction Network Database (BIND)—a standardized repository for all forms of biological interaction data, including protein–protein interactions²⁵. To systematically compile a set of published interactions, we used a search engine called PreBIND (J. Martin *et al.*, in preparation), which is a support vector machine and natural language-processing-based algorithm designed to identify abstracts that describe protein–protein interactions. Beginning with all 600 bait proteins used in this study, PreBIND was used to collect a non-exhaustive set of 697 protein interactions from the literature. The PreBIND data was combined with 545 interactions derived from the MIPS protein interaction table²⁶ to create a literature-based set of 1,003 interactions that involve the HMS-PCI bait set (see Supplementary Information). When compared against this literature benchmark, the HMS-PCI data set contained 2.6- to 3.4-fold more literature-derived interactions per bait than each large-scale HTP-Y2H data set, and 1.9-fold more interactions when compared with the combination of both comprehensive HTP-Y2H data sets (Fig. 3a, b and Table 1)^{3,4}. In addition to published interactions, a number of new interactions were shared by the HMS-PCI and HTP-Y2H data sets (Fig. 3c). Functional annotation of the HMS-PCI data set indicated that 275 of the detected complexes contained two or more interaction partners within the same gene ontology (GO) biological process (see Supplementary Information). Finally, we note that the connectivity distribution of the HMS-PCI data set follows a power law, as observed for other large-scale biological networks (see Supplementary Information Fig. 3 for a Pajek representation of the data set)²⁷.

Proteome-wide analysis of native protein complexes by highly sensitive mass spectrometric methods allows the detection of complex cellular networks that might otherwise elude more focused approaches (G.D.B. *et al.*, manuscript in preparation). Given that approximately 40% of yeast proteins are conserved through eukaryotic evolution²⁸, the global yeast protein interaction map will provide a partial framework for understanding more complex proteomes. Imminent technical advances, such as gel-free analysis

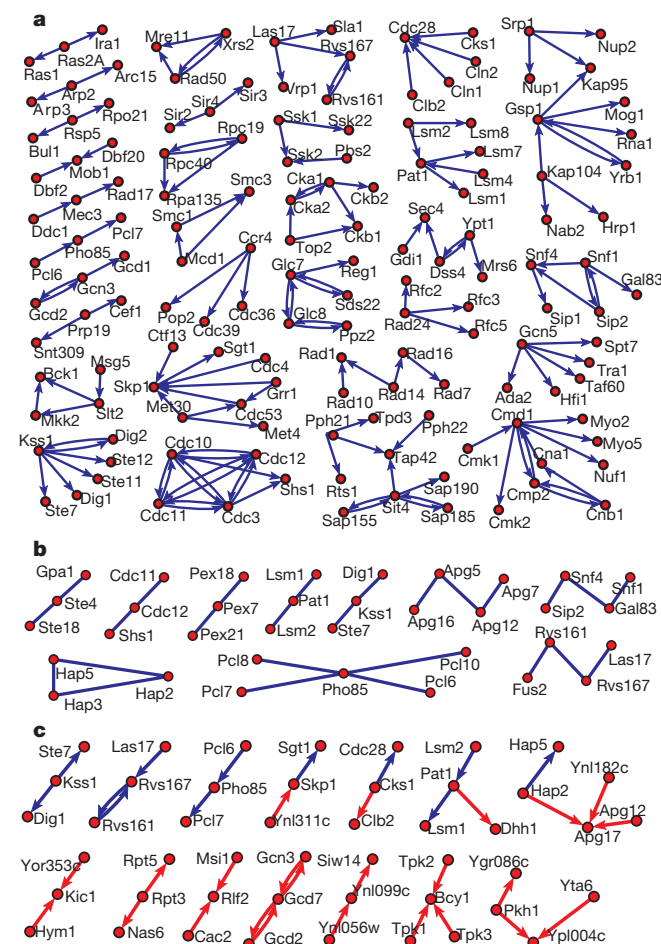


Figure 3 Comparison of large-scale protein interaction networks with interactions reported in the literature. **a**, Overlap of HMS-PCI data set and PreBIND + MIPS data set. **b**, Overlap of a comprehensive HTP-Y2H data set⁴ and PreBIND + MIPS data set. **c**, Overlap of HMS-PCI and the HTP-Y2H data set⁴. Blue edges are literature-derived interactions; red edges are new interactions detected by HTP approaches. For clarity, single binary interactions are not shown. The following number of single interactions were removed from the indicated panel: **a**, 37; **b**, 23; **c**, 31.

Table 1 Literature-derived interactions in HMS-PCI and HTP-Y2H interaction data sets

HTP data set	Literature data set*				
	PreBIND	MIPS	PreBIND+MIPS	MIPS two-hybrid	MIPS biochemical
HMS-PCI	113	119	166	55	81
Uetz	42	53	63	37	22
Ito full	37	39	49	27	18
Ito core	25	26	32	19	9
Ito full+Uetz	60	71	86	47	37

To address possible methodological bias in the literature benchmark, the MIPS data set was sorted into two-hybrid interactions (MIPS two hybrid) and interactions based on biochemical purification (MIPS biochemical). As these two sets are of roughly equal size, the MIPS benchmark is impartial in this aspect. Two large-scale HTP-Y2H data sets were used (Uetz, P. *et al.* (ref. 3) and Ito, T. *et al.* (ref. 4)). 'Ito full' refers to the complete set of 4549 interactions; 'Ito core' refers to the core set of the 841 most reproducible interactions, as defined by ref. 4.

*Number of interactions reported in the literature. See Supplementary Information for details on interaction curation and comparisons.

of protein complexes, higher sensitivity mass spectrometers, systematic analysis of post-translational modifications, and protein microarrays, will undoubtedly extend the reach of the approach described here^{6,29}. As the set of proteins nominally encoded by the human genome is approximately 5-fold greater than the total number of yeast proteins, comprehensive analysis of the human proteome is feasible with current technology. □

Methods

Recombination-based cloning, yeast culture and isolation of protein complexes were carried out using standard methods (see Supplementary Information for full details of methods). Cultures of strain BY4742 (*MAT α his3 Δ 1 leu2 Δ 0 lys2 Δ 0 ura3 Δ 0 pep4 Δ :: KANR*) bearing bait plasmids were induced by either addition of galactose or doxycyclin before collection at mid-logarithmic phase ($A_{600} = 1.2$ – 1.5). After immunoprecipitation of bait complexes, protein bands were visualized by colloidal Coomassie stain, excised from polyacrylamide gels, reduced and S-alkylated, then subjected to trypsin hydrolysis³⁰. To achieve high throughput, we constructed an automated proteomics network of mass spectrometers based on nano high-performance liquid chromatography (HPLC)-electrospray ionization-MS/MS capable of continuous operation. Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis was performed on a Finnigan LCQ Deca ion trap mass spectrometer (Thermo Finnigan) fitted with a Nanospray source (MDS Proteomics). Chromatographic separation was conducted using a Famos autosampler and an Ultimate gradient system (LC Packings) over Zorbax SB-C18 reverse phase resin (Agilent) packed into 75 μ m ID PicoFrit columns (New Objective). Protein identifications were made using the commercially available search engines Mascot (Matrix Sciences), Sonar (Proteometrics), Sequest (ThermoFinnigan) and PepSea (MDS Proteomics). Both the raw and filtered data sets generated in this study are available at <http://www.mdsf.com/yeast>. The filtered data set has been deposited in BIND²³ and can be viewed at <http://www.bind.ca/>.

Received 27 September; accepted 5 December 2001.

- Pawson, T. & Nash, P. Protein–protein interactions define specificity in signal transduction. *Genes Dev.* **14**, 1027–1047 (2000).
- Fields, S. & Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246 (1989).
- Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
- Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
- Neubauer, G. *et al.* Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl Acad. Sci. USA* **94**, 385–390 (1997).
- Mann, M., Hendrickson, R. C. & Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **10**, 437–473 (2001).
- Gustin, M. C., Albertyn, J., Alexander, M. & Davenport, K. MAP kinase pathways in the yeast *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **62**, 1264–1300 (1998).
- Morgan, D. O. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu. Rev. Cell. Dev. Biol.* **13**, 261–291 (1997).
- McMillan, J. N. *et al.* The morphogenesis checkpoint in *Saccharomyces cerevisiae*: cell cycle control of Swe1p degradation by Hsl1p and Hsl7p. *Mol. Cell. Biol.* **19**, 6929–6939 (1999).
- Philips, J. & Herskowitz, I. Identification of Klp1, a kelch domain-containing protein involved in cell fusion and morphology in *Saccharomyces cerevisiae*. *J. Cell. Biol.* **143**, 375–389 (1998).
- Jorgensen, P. & Tyers, M. The forked path to mitosis. *Genome Biol.* **1**, 1022.1–1022.4 (2000).
- Alexandru, G., Uhlmann, F., Mechtler, K., Poupard, M. & Nasmyth, K. Phosphorylation of the cohesin subunit Scc1 by Polo/Cdc5 kinase regulates sister chromatid separation in yeast. *Cell* **105**, 459–472 (2001).
- Zhou, B. B. & Elledge, S. J. The DNA damage response: putting checkpoints in perspective. *Nature* **408**, 433–439 (2000).
- Prakash, S. & Prakash, L. Nucleotide excision repair in yeast. *Mutat. Res.* **451**, 13–24 (2000).
- Thelen, M. P., Venclovas, C. & Fidelis, K. A sliding clamp model for the Rad1 family of cell cycle checkpoint proteins. *Cell* **96**, 769–770 (1999).
- Koegl, M. *et al.* A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. *Cell* **96**, 635–644 (1999).
- Ortolan, T. G. *et al.* The DNA repair protein Rad23 is a negative regulator of multi-ubiquitin chain assembly. *Nature Cell Biol.* **2**, 601–608 (2000).
- Tyres, M. & Rottapel, R. VHL: a very hip ligase. *Proc. Natl Acad. Sci. USA* **96**, 12230–12232 (1999).
- Emili, A., Schieltz, D. M., Yates, J. R. & Hartwell, L. H. Dynamic interaction of DNA damage checkpoint protein Rad53 with chromatin assembly factor Asf1. *Mol. Cell* **7**, 13–20 (2001).
- Hu, F., Alcasabas, A. A. & Elledge, S. J. Asf1 links Rad53 to control of chromatin assembly. *Genes Dev.* **15**, 1061–1066 (2001).
- Marsolier, M. C., Roussel, P., Leroy, C. & Mann, C. Involvement of the PP2C-like phosphatase Ptc2p in the DNA checkpoint pathways of *Saccharomyces cerevisiae*. *Genetics* **154**, 1523–1532 (2000).
- Durocher, D., Henckel, J., Fersht, A. R. & Jackson, S. P. The FHA domain is a modular phosphopeptide recognition motif. *Mol. Cell* **4**, 387–394 (1999).
- Zhao, X., Chabes, A., Domkin, V., Thelander, L. & Rothstein, R. The ribonucleotide reductase inhibitor Sml1 is a new target of the Mec1/Rad53 kinase cascade during growth and in response to DNA damage. *EMBO J.* **20**, 3544–3553 (2001).
- Beaudenon, S. L., Huacani, M. R., Wang, G., McDonnell, D. P. & Huijbregts, J. M. Rsp5 ubiquitin-protein ligase mediates DNA damage-induced degradation of the large subunit of RNA polymerase II in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19**, 6972–6979 (1999).
- Bader, G. *et al.* BIND—The biomolecular interaction network database. *Nucleic Acids Res.* **29**, 242–245 (2001).

- Mewes, H. W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **28**, 37–40 (2000).
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
- Chervitz, S. A. *et al.* Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* **282**, 2022–2028 (1998).
- Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105 (2001).
- Wilm, M. *et al.* Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466–469 (1996).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank J. Chen, B. Kuehl, H. Li, V. Lay, B. Tuekam, S. Zhang, M. Patel, P. O'Donnell, I. Dutschek, U. Friedrich, M. Hansen, J. Brønd, H. Lieu, R. Woolstencroft, L. Harrington, F. Sicheri, A. Breittkreutz, C. Boone, B. Andrews and T. Hughes for discussions and/or technical assistance. This work was supported in part by grants from the Canadian Institutes of Health Research (CIHR), the Ontario Research and Development Challenge Fund and MDS-Sciex to T.P., D.D., C.H. and M.T. T.P. is a Distinguished Scientist of the CIHR; M.F.M. is a CIHR Scientist; D.D. is a Canada Research Chair in Proteomics, Bioinformatics and Functional Genomics and a Hitchings-Elion fellow of the Burroughs-Wellcome Fund; and M.T. is a Canada Research Chair in Biochemistry.

Competing interests statement

The authors declare competing financial interests: details accompany the paper on Nature's website (<http://www.nature.com>).

Correspondence and requests for materials should be addressed to D.F. (e-mail: dfgeys@mdsp.com).

Alternative nucleotide incision repair pathway for oxidative DNA damage

Alexander A. Ischenko*† & Murat K. Saparbaev*

*Groupe “Réparation de l'ADN”, UMR 8532 CNRS, LBPA-ENS Cachan, Institut Gustave Roussy, 94805 Villejuif Cedex, France
†Novosibirsk Institute of Bioorganic Chemistry, Siberian Division of the Russian Academy of Sciences, 630090 Russia

The DNA glycosylase pathway¹, which requires the sequential action of two enzymes for the incision of DNA², presents a serious problem for the efficient repair of oxidative DNA damage, because it generates genotoxic intermediates such as abasic sites and/or blocking 3'-end groups that must be eliminated by additional steps before DNA repair synthesis can be initiated. Besides the logistical problems, biological evidence hints at the existence of an alternative repair pathway. Mutants of *Escherichia coli*³ and mice (ref. 4 and M. Takao *et al.*, personal communication) that are deficient in DNA glycosylases that remove oxidized bases are not sensitive to reactive oxygen species, and the *E. coli* triple mutant *nei*, *nth*, *fpg* is more radioresistant than the wild-type strain⁵. Here we show that Nfo-like endonucleases nick DNA on the 5' side of various oxidatively damaged bases, generating 3'-hydroxyl and 5'-phosphate termini. Nfo-like endonucleases function next to each of the modified bases that we tested, including 5,6-dihydrothymine, 5,6-dihydrothymine, 5-hydroxyuracil and 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine residues. The 3'-hydroxyl terminus provides the proper end for DNA repair synthesis; the dangling damaged nucleotide on the 5' side is then a good substrate for human flap-structure endonuclease⁶ and for DNA polymerase I of *E. coli*.

Because *E. coli* *nfo*[−] and *xth*[−] strains^{7,8} are extremely sensitive to oxidizing agents, we investigated whether the known abasic (AP) endonucleases, namely Nfo, Xth and human APEX/Ref-1/Ape1/

HAP-1 (ref. 9) proteins, can recognize oxidative DNA lesions that are already known to be excised by DNA glycosylases. The 5,6-dihydropyrimidines (DHPs) are the main products generated by γ -irradiation under anoxic conditions¹⁰; therefore, we tested the activity of Nfo, Xth and HAP-1 proteins on 30-nucleotide (30-mer) duplex oligonucleotides containing tetrahydrofuran (THF), 5-hydroxyuracil (5ohU), 5,6-dihydrouracil (DHU) and 5,6-dihydrothymine (DHT) residues at position 11 opposite adenine (A) for DHT or opposite guanine (G) for the other modified bases.

When the oligonucleotides containing DHPy are exposed to piperidine under conditions that preferentially cleave AP sites, they are barely incised. Thus, the amount of AP sites, which are the main substrates for AP endonucleases, is negligible (<3%; Fig. 1a, lanes 6–8). Nfo incises DHU·G, DHT·A and 5ohU·G duplex oligonucleotides to generate 10-mer fragments (Fig. 1a, lanes 10–12), but neither Xth nor HAP-1 (lanes 14–16 and 18–20) cleaves these substrates efficiently. In addition, Nfo, but neither Xth nor HAP-1, cleaves oligonucleotide adjacent to 2,6-diamino-4-hydroxy-5-*N*-methylformamidopyrimidine (Fapy) residues, when present as a duplex or in plasmid DNA (data not shown). These results indicate that, out of all the AP endonucleases tested, only Nfo can process a broad number of DNA lesions generated by oxidative stress.

Consistent with this, overproduction of the Nfo protein fully restores drug resistance to *E. coli* *xth*-deficient mutants, but the converse has not been observed¹¹. Furthermore, the Nfo protein hydrolytically incises DNA containing α -deoxyadenosine¹² and a benzene-derived exocyclic adduct¹³, which are not substrates for known DNA glycosylases, leaving the dangling nucleotide¹³ on the 5' end of the cleavage fragment. Indeed, the crystal structure of Nfo bound to oligonucleotide containing an AP site reveals a solvent-accessible pocket on the enzyme surface that could accommodate damaged bases¹⁴, providing structural basis for broad damage

specificity of Nfo. Together with these studies, our data indicate that Nfo may be a back-up enzyme to DNA glycosylases and may provide support for the existence of an alternative *E. coli* repair pathway for oxidative DNA adducts.

To determine the mechanism of incision of DHU by the Nfo protein, 3'-[³²P]ddAMP-labelled duplex oligonucleotides containing an AP site or DHU residues were generated. Nfo protein cleaves an AP site, derived from uracil (U) excised by uracil-DNA glycosylase (UDG), and generates a 20-mer DNA fragment lacking the U residue at the 5' end (Fig. 1b, lane 5), a result that is consistent with the established mechanism of action of DNA glycosylases¹⁵. Note that the deoxyribose 5'-phosphate (dRP) residue generated by Nfo acting on an AP site is usually lost during electrophoresis owing to its extreme lability¹⁶. When the dRP residue is stabilized by NaBH₄ treatment, however, the fragment migrates more slowly than an oligonucleotide of the same sequence, because of the reduced sugar moiety at the 5' end (lane 6).

Acting on DHU·G, Nfo generates two DNA fragments (lane 7). The faster migrating minor DNA fragment corresponds to the product observed in lanes 5 and 3, and is derived from an AP site. The major reaction product in lane 7 migrates slower than the product in lane 6 and does not change migration rate after piperidine treatment (lane 9), suggesting the presence of a non-excised DHU residue at the 5' end. When treated with calf intestinal phosphatase, DNA fragments generated by Nfo and Nth migrate slowly, indicating the presence of a 5'-phosphate (lanes 8 and 11). Together, the analysis of the mobility patterns of the cleavage products strongly indicates that Nfo acts in a completely different manner to the DNA glycosylase/ β -lyase Nth protein. By hydrolysing the phosphodiester bond 5' to the DHU residue, Nfo leaves the lesion attached to the 5' end of the downstream fragment and an OH group on the 3' end of the nicked site. These products generate, in a single step, a proper end for DNA repair synthesis.

To address further the role of Nfo in recognizing DHPy, we tested its activity in *E. coli* strains containing either the wild-type or a mutant (*nfo*⁻) Nfo gene (Fig. 2). When using 3'-[³²P]ddAMP-labelled oligonucleotides, both a DNA glycosylase and an endonuclease activity towards DHU were detected in all *E. coli* strains tested (Fig. 2a, lanes 7, 8, 10, 11), except the *nfo*⁻ strain (lane 9). The addition of paraquat to growing cultures of wild type (lane 8) greatly increased the endonuclease activity. As expected from the known induction of the Nfo protein by this drug¹⁷, we did not detect this activity in the paraquat-induced *nfo*⁻ strain, in contrast to wild-type, *fpg*⁻, *nth*⁻, *nei*⁻ strains (Supplementary Information Fig. 1). Finally, using the same oligonucleotide labelled at the 5' end with ³²P, we found that the *nfo*⁻ strain lacks the endonucleolytic activity that generates 3'-OH termini (Fig. 2b, lane 4; see Supplementary Information Fig. 2 for whole gel). These data show that the Nfo protein is the enzyme that directly incises the DHU in *E. coli* extracts.

Both the AP endonuclease and a DNA glycosylase can directly act on oxidative DNA lesions in *E. coli* extracts. Comparing the kinetic parameters for purified enzymes shows that Nth, a DNA glycosylase, cleaves the DHU·G duplex oligonucleotides more efficiently than Nfo, an AP endonuclease (Supplementary Information Table 1). But Nfo may be important *in vivo*, especially upon induction of the soxRS regulon by diverse environmental oxidative stresses¹⁸.

In search of an Nfo-like activity in eukaryotes, we examined the processing of 3'- and 5'-labelled oligonucleotides containing DHU, DHT, 7,8-dihydro-8-oxoguanine (OG) and U residues in whole-cell extracts from *Saccharomyces cerevisiae* and HeLa cells (Fig. 3). DNA fragments generated by purified enzymes acting on DHU and reduced AP (rAP) sites were used as size markers. Because the 3'-labelled fragments generated either by Fpg or Nth proteins on incision of DHU have phosphorylated 5' ends, their migration patterns are similar (Fig. 3a, lanes 5, 6). The migration pattern of incision products of DHPy in yeast extracts indicates that the Nfo-

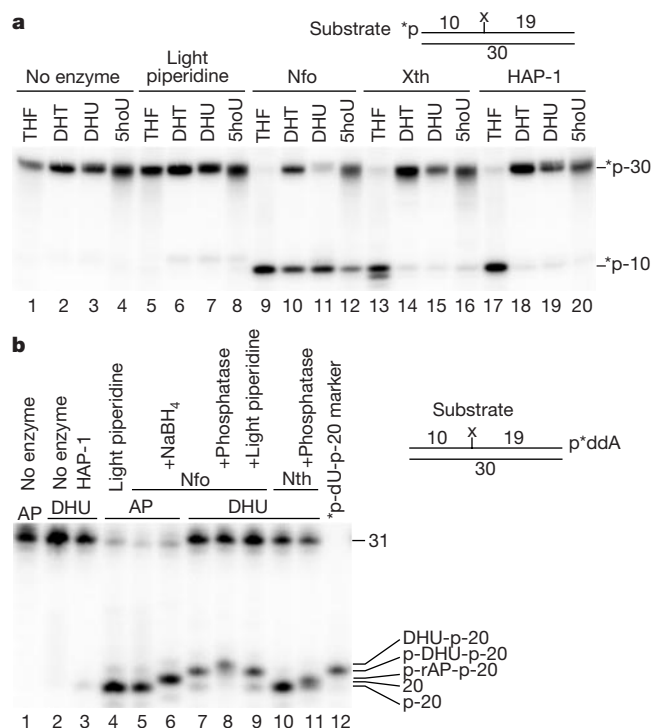


Figure 1 Incision activity of various AP endonucleases towards oxidative DNA lesions. **a**, 5'-³²P-labelled (*p) duplex oligonucleotides containing THF·G, DHU·G, DHT·A or 5ohU·G were incubated with excess amounts of Nfo, Xth and HAP-1 (60 ng). **b**, Determination of end groups generated by Nfo using 3'-[³²P]ddAMP-labelled (p*ddA) duplex oligonucleotides. rAP, reduced AP site.

like activity is the major activity against these lesions observed in *S. cerevisiae* (Fig. 3a, lanes 9, 10), whereas the DNA glycosylase (p-20 fragments) is the minor activity. In HeLa cell extracts, DHPy is mainly repaired by a DNA glycosylase, but the presence of a second reaction product (p-DHPy-p-20 fragments) strongly indicates the existence of an Nfo-like activity in human cells (Fig. 3a, lanes 12, 13). Thus, the presence of a significant endonuclease activity in yeast and HeLa cells, in addition to the DNA glycosylase, indicates that this pathway may be evolutionarily conserved from *E. coli* to humans.

Yeast extracts generate only DNA fragments containing a 3'-OH group during the repair of 5'-labelled oligonucleotides containing U (used as a control) and DHPy residues (Fig. 3b, lanes 8, 9, 11). This observation is consistent with cleavage mediated by an Nfo-like enzyme. Moreover, analysis of the reaction products generated with HeLa cells extracts suggests that there may be an Nfo-like activity in addition to DNA glycosylase/AP lyase activities on all substrates tested (Fig. 3b, lanes 12–15), although products containing 3'-OH groups might be due, in part, to the 3'-phosphoglycolate (PGA) activity present in HeLa cells (data not shown). Together, the results obtained with both 3'- and 5'-labelled substrates establish that yeast and human cells express a true endonucleolytic activity towards DHPy. In contrast to the yeast genome, however, the human genome apparently lacks an *nfo* homologue, suggesting that this activity in human cells is provided by a separate, as yet unidentified, enzyme.

The yeast *Apn1* protein is homologous to *Nfo*¹⁹. We therefore investigated the repair of 3'-labelled DHU-G or U-G duplex oligonucleotides in *S. cerevisiae* strains deficient in *Apn1* and/or *Apn2* endonucleases. Incision of the DHU-G oligonucleotide in yeast cell extracts, in the absence of Mg^{2+} , had an absolute requirement for the *apn1*, but not the *apn2* gene product (Fig. 3c, lanes 4–7), whereas cleavage of the U-G-containing oligonucleotide was observed in all mutants tested (lanes 8–11). Similar results were obtained when yeast cell extract was supplemented with 5 mM Mg^{2+} (Supplementary

Information Fig. 3). These observations indicate that in *S. cerevisiae* *Apn1* is the principal enzyme that repairs DHPy, and it seems to be more efficient than its bacterial counterpart. It is tempting to propose that the increased spontaneous mutation rate observed in the *S. cerevisiae apn1* mutants is, in part, caused by the accumulation of unrepaired modified bases in addition to AP sites.

Depending on the organism, the excision repair of a given ultraviolet lesion is initiated by different pathways including DNA glycosylase¹⁵, nucleotide excision repair²⁰ and alternative excision repair (UV-endonuclease)²¹. We propose that a similar set of pathways exists for the excision repair of oxidative DNA damage, including DNA glycosylase (refs 3–5 and M. Takao *et al.*, personal communication), nucleotide excision repair²² and AP endonuclease (our results).

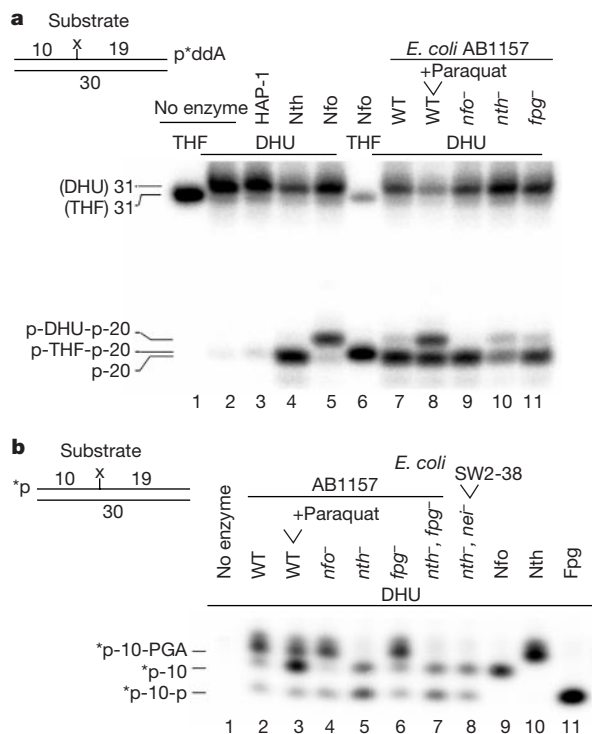


Figure 2 *Nfo* protein is the endonuclease that recognizes DHPy. The 3'-[³²P]ddAMP- (a) or 5'-[³²P]- (b) labelled DHU-G duplex oligonucleotides were incubated in cell-free extracts from *E. coli*. PGA, phosphoglycolate.

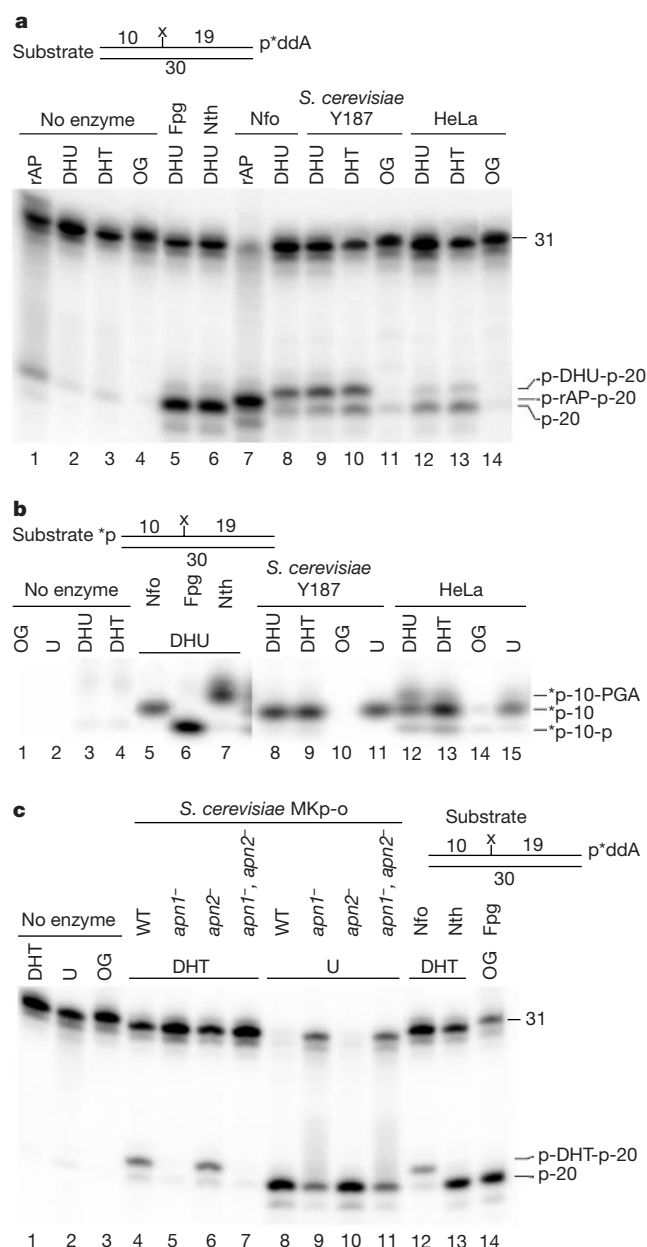


Figure 3 Cleavage of oligonucleotides containing modified bases by whole-cell extracts from *S. cerevisiae* and HeLa cells. **a**, Assays were carried out with a 3'-[³²P]ddAMP-labelled duplex oligonucleotides containing DHU-G, DHT-A, OG-C and rAP-G. **b**, As in **a** but using 5'-[³²P]-labelled DHU-G, DHT-A, OG-C and U-G. **c**, The 3'-[³²P]ddAMP-labelled DHT-G or U-G duplex oligonucleotides were incubated in cell-free extracts from *S. cerevisiae* wild-type and mutant strains.

Base excision repair proceeds through two alternative mechanisms dependent on either a 'short-patch' DNA polymerase β (pol β) pathway or a 'long-patch' flap-structure endonuclease (FEN-1) pathway²³. In the former, pol β removes through β -elimination a 5'-dRP moiety, left after the incision of unmodified natural AP sites by HAP-1. Consequently, repair patches 2–6 nucleotides long (long-patch) are found after repair of a reduced or oxidized abasic sites²⁴. Furthermore, the same observation is true for the UV-damage endonuclease (UVDE)-initiated repair pathway that has been reconstituted *in vitro* using the FEN-1 equivalent, Rad2 (ref. 25). As the dangling nucleotide generated by Nfo/Apn1-like enzyme cannot be simply removed by β -elimination, we predicted

that the DNA glycosylase-independent repair of DHPy occurs through the long-patch pathway.

To establish whether an 5'-exonuclease activity could remove the dangling nucleotide, we prepared a substrate in which a 44-mer was annealed to a 5'-labelled 21-mer containing a DHU residue at its 5' end and a 19-mer oligonucleotide (Fig. 4a). This construction yields a duplex DNA that is completely nicked 5' to the DHU to produce a product identical to the one generated by Nfo protein. Pol I efficiently eliminates the 5'-terminal DHU residue from the nicked duplex oligonucleotide as a mononucleotide (Fig. 4a, lane 3). Addition of dNTPs to the reaction enhance removal of the DHU from DNA (lane 4). FEN-1 eliminates DHU poorly (lane 5); however, addition of human proliferating cell nuclear antigen (PCNA) greatly stimulates DHU removal (lane 6). These results establish that pol I has a role in *E. coli* and FEN-1/PCNA in human cells in the efficient elimination of a dangling modified nucleotide from the 5' end, despite the fact that it does not form a stable flap structure.

The DNA glycosylase-independent incision of oxidatively damaged DNA by Nfo/Apn1-like enzymes provides an alternative pathway to classic base excision repair (Fig. 4b). We propose to name it the 'nucleotide incision repair' pathway. This pathway has the advantage of avoiding the genotoxic intermediates generated in the base excision repair pathway. Nucleotide incision repair provides an explanation for the DNA repair proficiency of DNA glycosylase-deficient mutants and supports the existence of a back-up repair pathway in *E. coli*, yeast and humans; moreover, it provides a new physiological target—a modified base rather than an artificial reduced abasic site—for the long-patch repair pathway described in human cells²⁴. Further experiments are required to fully investigate this process. □

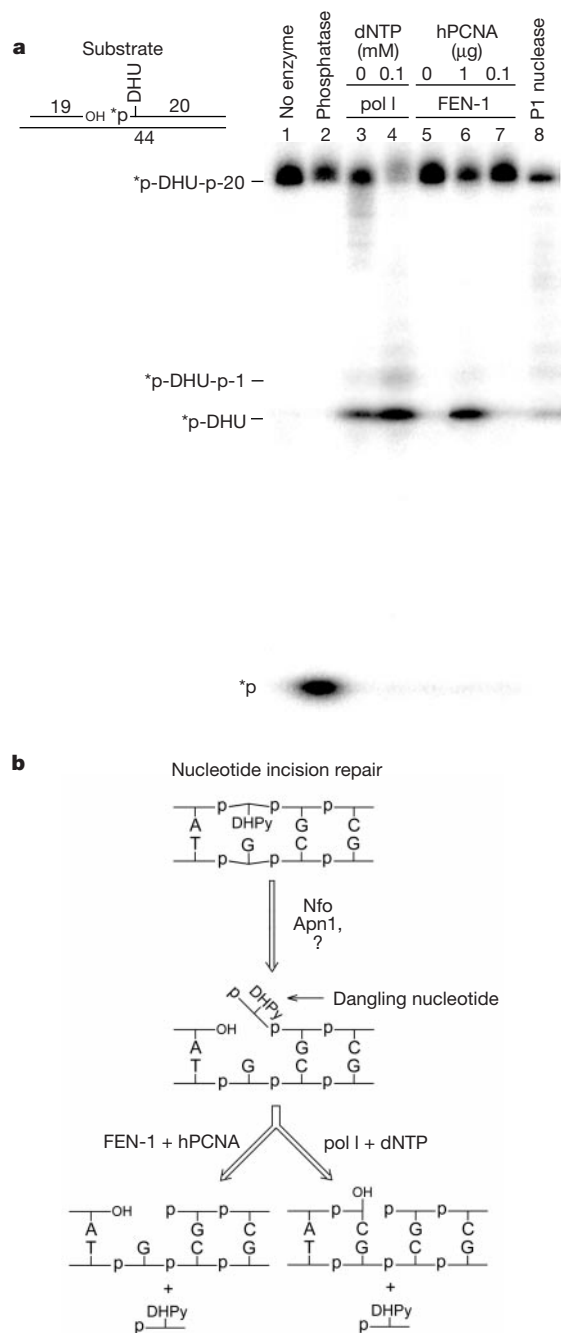


Figure 4 Exonucleolytic activities of *E. coli* polymerase I and human FEN-1 proteins at the nicked duplex DNA containing 5'-dangling nucleotide. **a**, Analysis of the products of the exonuclease activities of pol I (0.08 units) and FEN-1 (32 ng) proteins at nicks containing 5'-³²P-labelled DHU. **b**, Production and elimination of a dangling modified nucleotide in the nucleotide incision repair pathway.

Methods

Oligonucleotides and chemicals

All oligodeoxyribonucleotides were purchased from Eurogentec, including the following modified oligonucleotides containing either a DHU, DHT, 5ohU, OG, U or THF residue: 5'-TGACTGCATAXGCATGTAGACGATGTGCAT-3' (30-mer), 5'-XCGCTGGTACC-CATCTCATGA-3' (21-mer), 5'-UGCATGTAGACGATGTGCAT-3' (21-mer), where X is the position of the modified base; complementary oligonucleotides, containing either dA, dG, dC or T opposite to the modified base; and oligonucleotides used for assembly of the nicked structure: 5'-AATTGCTATCTAGCTCATA-3' (19-mer), 5'-CTTCATGAGATGGG TACCAGCGGTATGAGCTAGATAGCAATTCA-3' (44-mer). We purchased Complete EDTA free (protease inhibitor cocktail), dNTPs and other molecular biology products from Roche.

End labelling

Oligonucleotides were either 5'-end labelled by T4 polynucleotide kinase (New England Biolabs) in the presence of [γ -³²P]ATP (4,500 Ci mmol⁻¹; ICN Pharmaceuticals), or 3'-end labelled by terminal transferase (New England Biolabs) in presence of [α -³²P]ddATP (5,000 Ci mmol⁻¹; Amersham-Pharmacia) as recommended by the manufacturers.

Enzymes and strains

We purified *E. coli* FPG, Nfo, UDG and Nth²⁶ and human HAP-1 (ref. 27) and FEN-1 (ref. 6) proteins as described. The AP endonuclease preparations lacked detectable AP lyase and DNA glycosylase activities on the substrates containing AP sites, 8-oxoG, U, alkylpurines and other modified bases (data not shown). Human PCNA protein was provided by G. Baldacci. *Penicillium citrinum* nuclease P1 was from Amersham-Pharmacia, and bovine serum albumin (BSA), calf intestinal alkaline phosphatase, Xth and pol I from Roche. *E. coli* strain AB 1157 (wild-type) and its isogenic derivatives, BH20 (*fpg::kan^R*), BH150 (*nth::kan^R*), BH130 (*nfo::kan^R*) and BH160 (*nth::kan^R*, *fpg::kan^R*, *X::tet^R*) were from our laboratory; and SW2-38 (isogenic to BW35 (KL16) except *nth::kan^R*, *nei::kan^R*) was a gift from S. Wallace. We purchased *S. cerevisiae* strain Y187 (MAT α , *ura3-52*, *his3-200*, *ade2-101*, *trp1-901*, *leu2-3*, *112*, *gal4Δ*, *gal80Δ*, *met⁻*, *URA3::GAL1 UAS-GAL1 TATA-lacZ*, MEL1) from Clontech; strains isogenic to MKP-o (MAT α , *leu2-3*, *112*, *his3-Δ1*, *trp1-289am*, *ura3-52*, *gal2*), DRY373 (*apn1Δ::his3*), RBY31 (*apn2Δ::hisG*), RBY71 (*apn1Δ::his3*, *apn2Δ::hisG*) were provided by D. Ramotar.

Preparation of whole cell-free extracts

Escherichia coli cell-free extracts were prepared from cultures grown to an absorbance of 1.0 at 600 nm as described²⁸, except that the lysis buffer contained 0.1 M HEPES (pH 7.5), 0.1 mM EDTA and protease inhibitors. Treatment by paraquat was performed as described¹⁷. We prepared *S. cerevisiae*¹⁹ and HeLa²⁹ whole cell-free extracts as described,

except that the lysis buffer contained 50 mM HEPES (pH 7.5), 0.1 mM EDTA, 0.5 M KCl, 5 mM β -mercaptoethanol, 0.1% NP-40, 5% glycerol and protease inhibitor cocktail (Roche).

Enzyme assays

To determine endonuclease activity, 1 pmol of 5'-³²P- or 3'-[³²P]ddAMP-labelled oligonucleotide duplex was incubated with either 6 ng of the given repair protein(s) for 20 min or 6 μ g of cell-free extract for 90 min (or 30 min for *E. coli*) at 37 °C in reaction buffer (20 μ l) containing 20 mM HEPES (pH 7.5), 50 mM KCl, 0.1 mM EDTA (or 2 mM EDTA for the extracts or 5 mM MgCl₂ for the FEN-1, HAP-1 and pol I proteins or 5 mM CaCl₂ for the Xth protein), 1 mM β -mercaptoethanol and 0.1 mg ml⁻¹ BSA, unless otherwise specified. When necessary, we performed light piperidine treatment by adding 10% piperidine for 15 min at 37 °C. Reaction products were analysed by electrophoresis on denaturing 20% polyacrylamide gels (7 M urea, 0.5 $\times$ TBE), visualized by PhosphorImager Storm 840 (Molecular Dynamics) and quantified using ImageQuant software.

Received 9 July; accepted 23 October 2001.

1. Lindahl, T. New class of enzymes acting on damaged DNA. *Nature* **259**, 64–66 (1976).
2. Laval, J. Two enzymes are required for strand incision in repair of alkylated DNA. *Nature* **269**, 829–832 (1977).
3. Laval, J., Jurado, J., Saporito, S. & Sidorkina, O. Antimutagenic role of base-excision repair enzymes upon free radical-induced DNA damage. *Mutat. Res.* **402**, 93–102 (1998).
4. Klungland, A. *et al.* Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage. *Proc. Natl Acad. Sci. USA* **96**, 13300–13305 (1999).
5. Blaisdell, J. O. & Wallace, S. S. Abortive base-excision repair of radiation induced clustered DNA lesions in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **98**, 7426–7430 (2001).
6. Kim, K., Biade, S. & Matsumoto, Y. Involvement of flap endonuclease 1 in base excision DNA repair. *J. Biol. Chem.* **273**, 8842–8848 (1998).
7. Cunningham, R. P., Saporito, S., Spitzer, S. G. & Weiss, B. Endonuclease IV (*nfo*) mutant of *Escherichia coli*. *J. Bacteriol.* **168**, 1120–1127 (1986).
8. Rogers, S. G. & Weiss, B. Cloning of the exonuclease III gene of *Escherichia coli*. *Gene* **11**, 187–195 (1980).
9. Dimple, B., Herman, T. & Chen, D. S. Cloning and expression of APE, the cDNA encoding the major human apurinic endonuclease: definition of a family of DNA repair enzymes. *Proc. Natl Acad. Sci. USA* **88**, 10450–10454 (1991).
10. Furlong, E. A., Jorgensen, T. J. & Henner, W. D. Production of dihydrothymidine stereoisomers in DNA by γ -irradiation. *Biochemistry* **25**, 4344–4349 (1986).
11. Ramotar, D., Popoff, S. C. & Dimple, B. Complementation of DNA repair-deficient *Escherichia coli* by the yeast Apn1 apurinic/aprimidinic endonuclease gene. *Mol. Microbiol.* **5**, 149–155 (1991).
12. Ide, H., *et al.* Alpha-deoxyadenosine, a major anoxic radiolysis product of adenine in DNA, is a substrate for *Escherichia coli* endonuclease IV. *Biochemistry* **33**, 7842–7847 (1994).
13. Hang, B., Chenna, A., Fraenkel-Conrat, H. & Singer, B. An unusual mechanism for the major human apurinic/aprimidinic (AP) endonuclease involving 5' cleavage of DNA containing a benzene-derived exocyclic adduct in the absence of an AP site. *Proc. Natl Acad. Sci. USA* **93**, 13737–13741 (1996).
14. Hosfield, D. J., Guan, Y., Haas, B. J., Cunningham, R. P. & Tainer, J. A. Structure of the DNA repair enzyme endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell* **98**, 397–408 (1999).
15. McCullough, A. K., Dodson, M. L. & Lloyd, R. S. Initiation of base excision repair: glycosylase mechanisms and structures. *Annu. Rev. Biochem.* **68**, 255–285 (1999).
16. Bailly, V. & Verly, W. G. The multiple activities of *Escherichia coli* endonuclease IV and the extreme lability of 5'-terminal base-free deoxyribose 5-phosphates. *Biochem. J.* **259**, 761–768 (1989).
17. Chan, E. & Weiss, B. Endonuclease IV of *Escherichia coli* is induced by paraquat. *Proc. Natl Acad. Sci. USA* **84**, 3189–3193 (1987).
18. Gonzalez-Flecha, B. & Dimple, B. Genetic responses to free radicals. Homeostasis and gene control. *Ann. NY Acad. Sci.* **899**, 69–87 (2000).
19. Popoff, S. C., Spira, A. L., Johnson, A. W. & Dimple, B. Yeast structural gene (APN1) for the major apurinic endonuclease: homology to *Escherichia coli* endonuclease IV. *Proc. Natl Acad. Sci. USA* **87**, 4193–4197 (1990).
20. Petit, C. & Sancar, A. Nucleotide excision repair: from *E. coli* to man. *Biochimie* **81**, 15–25 (1999).
21. Yasui, A. & McCready, S. J. Alternative repair pathways for UV-induced DNA damage. *BioEssays* **20**, 291–297 (1998).
22. Reardon, J. T., Besho, T., Kung, H. C., Bolton, P. H. & Sancar, A. *In vitro* repair of oxidative DNA damage by human nucleotide excision repair system: Possible explanation for neurodegeneration in Xeroderma pigmentosum patients. *Proc. Natl Acad. Sci. USA* **94**, 9463–9468 (1997).
23. Wilson, D. M. III & Thompson, L. H. Life without DNA repair. *Proc. Natl Acad. Sci. USA* **94**, 12754–12757 (1997).
24. Klungland, A. & Lindahl, T. Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1). *EMBO J.* **16**, 3341–3348 (1997).
25. Allea, J. L., Zuo, S. J., Hurwitz, J., & Doetsch, P. W. *In vitro* reconstitution of the *Schizosaccharomyces pombe* alternative excision repair pathway. *Biochemistry* **39**, 2659–2666 (2000).
26. Saporito, S. & Laval, J. 3,N⁴-ethenocytosine, a highly mutagenic adduct, is a primary substrate for *Escherichia coli* double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase. *Proc. Natl Acad. Sci. USA* **95**, 8508–8513 (1998).
27. Barzilay, G. *et al.* Identification of critical active-site residues in the multifunctional human DNA repair enzyme HAP1. *Nature Struct. Biol.* **2**, 561–568 (1995).
28. Boiteux, S., O'Connor, T. R. & Laval, J. Formamidopyrimidine-DNA glycosylase of *Escherichia coli*: cloning and sequencing of the *fpg* structural gene and overproduction of the protein. *EMBO J.* **6**, 3177–3183 (1987).
29. Maksimenko, A. V., Gottikh, M. B., Helin, V., Shabarova, Z. A. & Malvy, C. Physico-chemical and biological properties of antisense phosphodiester oligonucleotides with various secondary structures. *Nucleos. Nucleot.* **18**, 2071–2091 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank J. Laval for stimulating discussions and financial support; G. Baldacci for human PCNA; C. V. Privezentzev for human HAP-1 and FEN-1; A. Maksimenko and C. Delaporte for human cells; D. Ramotar for yeast strains; and P. R. Strauss, A. Kuzminov and D. Zharkov for critically reading the manuscript. This work was supported by the European Community (M.K.S.), Association pour la Recherche sur le Cancer (J.L.), and fellowships from Centre français pour l'accueil et les échanges internationaux (EGIDE) to A.A.I.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.K.S. (e-mail: smurat@igr.fr).

Co-regulator recruitment and the mechanism of retinoic acid receptor synergy

Pierre Germain, Jaya Iyer*, Christina Zechel* & Hinrich Gronemeyer

Institut de Génétique et de Biologie Moléculaire et Cellulaire, Centre National de la Recherche Scientifique, Institut National de la Santé et de la Recherche Médicale, Université Louis Pasteur, BP 163, F-67404 Illkirch Cedex, C. U. de Strasbourg, France

Crystal structure and co-regulator interaction studies have led to a general mechanistic view of the initial steps of nuclear receptor (NR) action. Agonist-induced transconformation of the ligand-binding domain (holo-LBD) leads to the formation of co-activator complexes, and destabilizes the co-repressor complexes bound to the ligand-free (apo) LBD^{1–3}. However, the molecular basis of retinoid-X receptor (RXR) 'subordination' in heterodimers, an essential mechanism to avoid signalling pathway promiscuity, has remained elusive. RXR, in contrast to its heterodimer partner, cannot autonomously induce transcription on binding of cognate agonists^{4–7}. Here we show that RXR can bind ligand and recruit co-activators as a heterodimer with apo-retinoic-acid receptor (apo-RAR). However, in the usual cellular environment co-repressors do not dissociate and they prohibit co-activator access because co-regulator binding is mutually exclusive. Accordingly, RXR subordination can be overcome in heterodimers that bind co-repressor weakly or in cells with a high co-activator content. We identify two types of RAR antagonists that differentially modulate co-regulator interaction, and we demonstrate that synergy between RAR ligands and RXR agonists^{6,8} results from increased interaction efficiency of a single p160 with the heterodimer, requiring two intact receptor-binding surfaces on the co-activator.

Electrophoretic mobility shift assays (EMSAs) demonstrated that RAR α Δ AB–RXR α Δ AB heterodimers (RAR–RXR; both heterodimer partners with a deleted AB region) form complexes with the NR-interacting domain (NID; see Fig. 1a and ref. 9) of the co-activator TIF2 in the presence of either RAR α agonists (Am80, BMS753) or the agonist 9-*cis* retinoic acid (9c-RA), which binds to both RARs and RXRs; however, this was not the case when the

* Present addresses: Monsanto Enterprises Limited, INBRI, 277 Chowdaiah Road, 18th cross Malleshwaram, Bangalore 560003, India (J.I.); Laboratory of Molecular Neurobiology, Institute of Biochemistry and Pathobiology, Johannes Gutenberg University Medical School, 55099 Mainz, Germany (C.Z.).

heterodimers were bound to a retinoic-acid receptor- α (RAR α)-selective (BMS614) antagonist (Fig. 1b). SR11237, an RXR-selective (referred to hereafter as rexinoid) agonist^{6,10}, induced formation of a ternary apo-RAR/holo-RXR/TIF2.42 complex on a cognate DR5 response element that migrated with the same mobility as the corresponding complex induced by RAR agonists. Electrospray ionization mass spectrometry (ESI-MS) with purified RAR–RXR heterodimer confirmed that pure rexinoids, such as CD3254, can bind efficiently to the RXR subunit in the absence of RAR ligands (Fig. 1c). Thus, binding of a rexinoid agonist to apo-RAR/RXR induces p160 co-activator association. We derived identical conclusions from experiments in which TIF2.42 was crosslinked to purified RAR–RXR heterodimers as TIF2.42 was recruited in the presence of both RAR and RXR agonists (Fig. 1d, top panel).

The interaction between the TIF2 NID and RAR–RXR could be mediated by one or both of the heterodimer partners; an alternative explanation is that TIF2 is bound independently by each hetero-

dimer partner. Disabling the activation function-2 (AF-2) by deleting helix 12 of RXR (generating a dominant negative RXR, dnRXR in Fig. 1b) impaired the ability of the corresponding heterodimer to recruit TIF2 in the presence of the RXR agonist SR11237; however, this did not affect TIF2 recruitment by RAR agonists. Of note, when RAR helix H12 was deleted (dnRAR), TIF2.42 bound to dnRAR–RXR in the presence of SR11237 (Fig. 1c, e, f) and 9c-RA (albeit less efficiently than to RAR–RXR), but not in the presence of RAR α agonists. These results were fully supported by crosslinking data using heterodimers purified to homogeneity (Fig. 1d, middle and bottom panels), as well as by EMSAs with TIF2.1 (Fig. 1e)—a functional co-activating TIF2 fragment¹¹ (Fig. 1a). TIF2.42M13 (containing one single NR box; Fig. 1a) bound efficiently to dnRAR–RXR in the presence of the rexinoid (Fig. 1f) but not Am80, thus excluding the possibility that dnRAR could possess a weak co-activator-binding surface that could synergize with a similar surface on RXR.

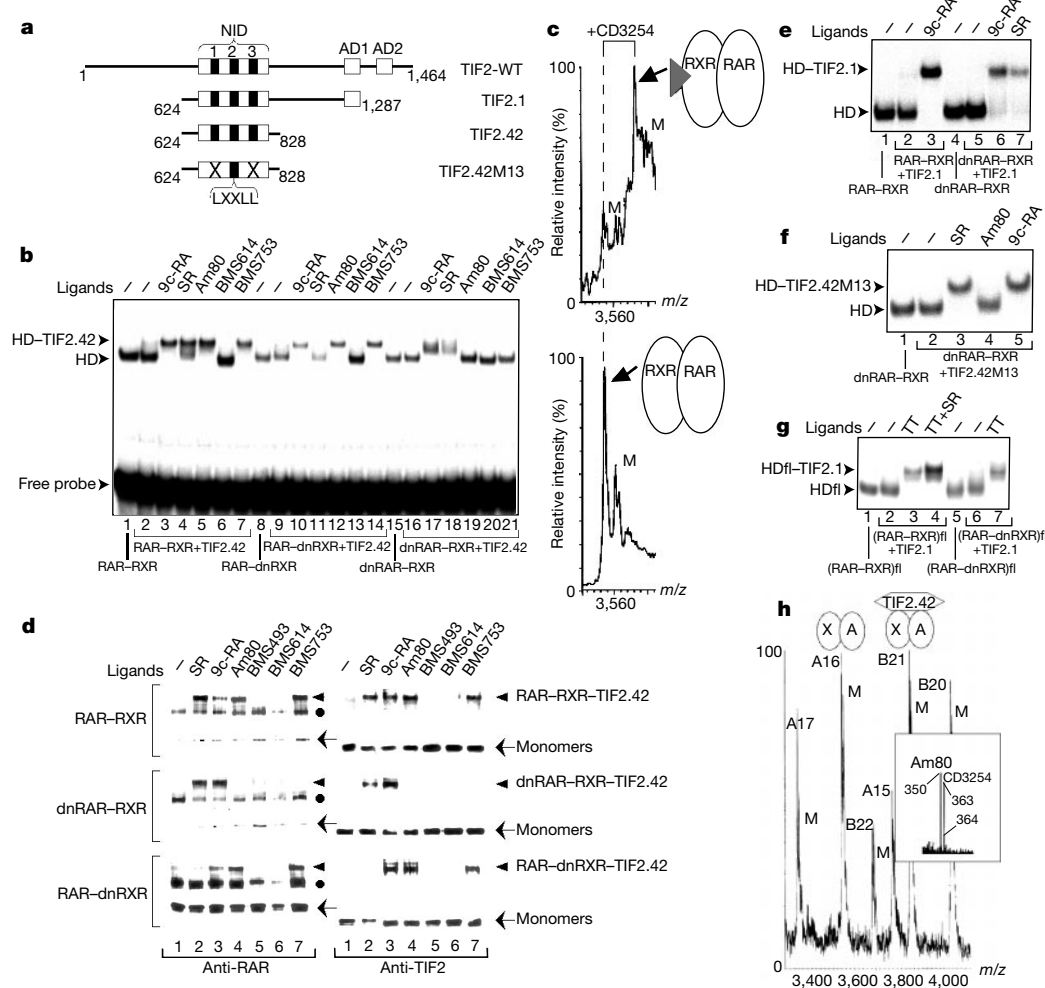


Figure 1 Both subunits of the RAR–RXR heterodimer can interact autonomously with a single TIF2 molecule. **a**, TIF2 constructs. Black rectangles represent three NR boxes within the NID; mutant TIF2.42M13 harbours only one functional NR box. TIF2-WT, wild-type TIF2. **b**, EMSAs demonstrating autonomous agonist and AF-2-dependent TIF2 recruitment by each subunit of the heterodimer (HD). SR, SR11237. **c**, Heterodimeric RXR autonomously binds cognate ligands. ESI-MS of purified ligand-free (bottom) and rexinoid CD3254-bound (top) RAR α –RXR α LBD heterodimers. The difference in mass between the two species (arrows; the 16 protonations area is shown) corresponds exactly to one molecule of CD3254 per heterodimer. M indicates complexes with N-terminal methionine in RXR. **d**, Purified heterodimers crosslinked with TIF2.42 in the presence of various

retinoids. Shown are western blots probed with anti-RAR or anti-TIF2 antibodies. Monomers indicate RAR, RXR and TIF2 monomers; filled circles indicate RAR–RXR heterodimers; arrowheads indicate the heterodimer–TIF2.42 complexes. **e–g**, Ligand-dependent formation of TIF2–heterodimer (HD) complexes in EMSAs with TIF2.1 (**e**), TIF2.42M13 (**f**), or full-length (fl) receptors (**g**) as indicated. **h**, ESI-MS (see Methods) of purified heterodimer (X–A)–TIF2.42 complexes containing both RAR (Am80) and RXR (CD3254) agonists reveals a single, bound TIF2.42 molecule. Inset, ESI-MS spectrum after denaturation of the same complex preparation confirming the presence of both RAR (Am80) and RXR (CD3254) ligands.

Multiple biochemical and structural data^{1,2} demonstrate the contribution of holo-LBD H12 to the co-activator-binding surface. Therefore, the above results provide strong evidence that both RAR and RXR subunits of the heterodimer not only autonomously bind their cognate ligands, but also that on agonist binding each of them can recruit p160 co-activators through a single surface comprising holo-H12. Our results do not support a previous model that allows RXR ligand binding and co-activator interaction only when RAR is bound to an agonist¹².

All agonist-induced TIF2.42 complexes of wild-type and dominant negative mutant receptors migrated in EMSAs with a mobility indistinguishable from that observed with TIF2.42M13. This same migration pattern (Fig. 1b), obtained in the presence of the RAR and RXR agonist 9c-RA, was surprising as the wild-type heterodimer could be expected to bind two TIF2 molecules, whereas heterodimers with one dominant negative receptor would recruit only one. At no time did we observe a band that corresponded to a heterodimer complex with two TIF2.42 molecules, and identical migrations were observed for the complexes between TIF2.1 and heterodimers comprising RAR α and full-length or dnRXR α (Fig. 1g). This suggests that a single TIF2 co-activator interacts with the heterodimer. Indeed, ESI-MS of the complex between TIF2.42 and the purified double-liganded RAR–RXR LBD heterodimer revealed only a single TIF2.42 molecule per heterodimer (Fig. 1h). Mass spectrometry showed equimolar amounts of Am80 and CD3254 bound to the heterodimer (Fig. 1h, inset). Thus, a single TIF2 molecule interacts with the holo-RAR/olo-RXR heterodimer, indicating that each subunit binds one of the three p160 NR boxes to generate a ternary complex.

RAR and RXR agonist synergy^{6,8} may result from cooperative p160 binding. As one co-activator molecule is bound per hetero-

dimer, cooperativity would originate from the simultaneous establishment of two heterodimer–co-activator interfaces, involving at least two NR boxes. Indeed, EMSAs with RAR–RXR heterodimers showed a strong synergy of RAR (TTNPB) and RXR (SR11237) agonists for TIF2.42 binding (Fig. 2a and data not shown). No synergy was seen with TIF2.42M13, which could establish only a single heterodimer–TIF2 interface. TIF2.42 mutants with singly disabled NR boxes (M1, M2, M3) revealed that two boxes, but not one (Fig. 2a and data not shown), suffice to generate RAR–RXR ligand synergy (Fig. 2b), which is most obvious for the weak⁹ NR boxes 1 and 3.

Cooperative TIF2 binding was also observed with some RAR antagonists. Together with the antagonist AGN (AGN192870; ref. 13) the rexinoid SR11237 induced formation of heterodimer–TIF2.42 complexes more efficiently than either ligand alone (Fig. 2c). Not considering RXR subordination, the ability of these compounds/combinations to promote heterodimer–TIF2.42 interaction parallels their ability to transactivate through a classical heterodimer–DR5-mediated pathway (Fig. 2d). Synergy between the RAR antagonist AGN and the pure RXR agonist CD3254 was further revealed by chromatin immunoprecipitation (ChIP) assays of the RAR β 2 gene promoter in NB4 promyelocytic leukaemia cells (Fig. 2e). Together with retinoid antagonists, pure rexinoid agonists can regulate transcription of endogenous target genes through endogenous receptors involving histone acetylation by recruitment of p160 co-activator complexes.

Both RAR and RXR ligands induce TIF2 recruitment; however, this creates a problem because RXR is ‘silenced’ by apo-RAR in heterodimers^{4–7} (Fig. 2d). We investigated whether co-repressor interaction could account for the inability of RXR to respond to rexinoid agonists in apo-RAR/RXR heterodimers. Silencing

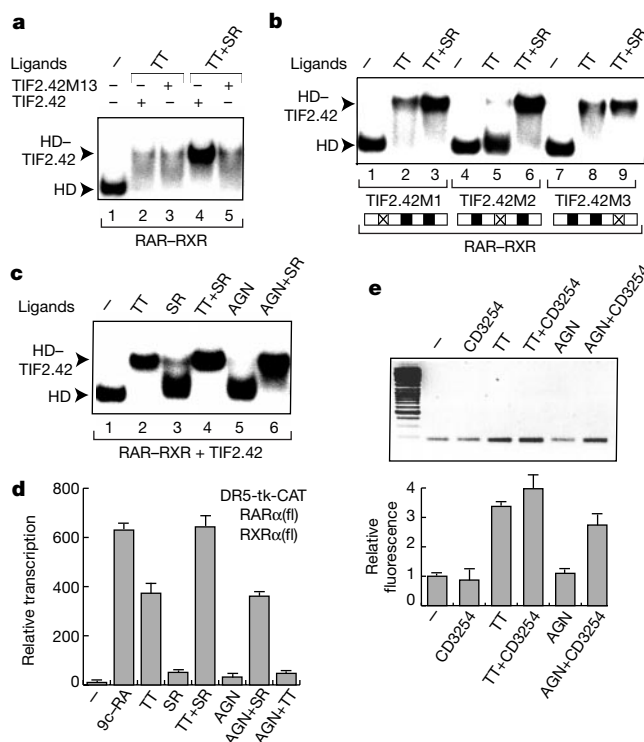


Figure 2 Cooperative recruitment of a single TIF2 molecule to RAR–RXR requires at least two intact NR boxes. **a**, EMSAs with TIF2.42 and heterodimers as indicated. TTNPB (TT; RAR agonist) and SR11237 (SR; rexinoid) cooperate strongly for recruiting TIF2.42 to RAR–RXR; no cooperativity occurs with TIF2.42M13. **b**, As in **a** but using TIF2 NR-box mutants (shown below). **c**, Cooperative TIF2.42 recruitment by heterodimers comprising the RAR antagonist AGN192870 (AGN)-bound and rexinoid agonist-bound subunits.

d, Transient transactivation in COS cells cotransfected with DR5-tk-CAT (ref. 30), and RAR α and RXR α full-length (fl) expression vectors reveals synergy between AGN and SR11237. **e**, ChIP assays with NB4 cells. Together CD3254 (rexinoid agonist) and AGN (retinoid antagonist) hyperacetylate histone H3 of chromatin encompassing the retinoic acid-responsive RAR β 2 promoter. Top, relative acetylation levels; bottom, quantitative acetylation levels (see Methods). Error bars in **d** and **e** indicate s.e.m.

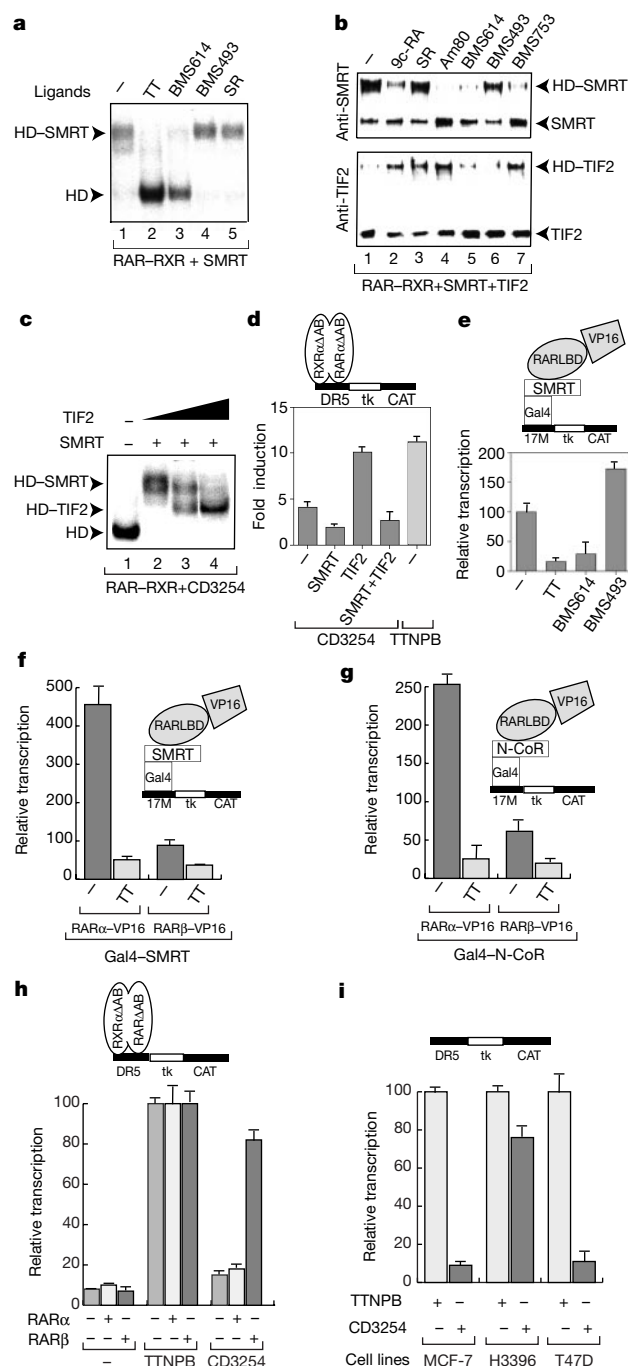


Figure 3 RAR-RXR heterodimer can interact with TIF2 or SMRT in the presence of the retinoid agonists, whereas the RAR antagonists impair TIF2 interaction and can destabilize (BMS614) or stabilize (BMS493) the SMRT-heterodimer interaction. **a**, EMSAs to assess SMRT interaction with heterodimers bound to the indicated ligands. **b**, Purified RAR-RXR heterodimers were crosslinked in the presence of the indicated ligands TIF2.42 and SMRT. **c**, Retinoid-bound heterodimer-SMRT complexes dissociate in favour of heterodimer-TIF2 complexes with increasing concentration of TIF2. **d**, Transient transactivation of DR5-tk-CAT with heterodimers in the presence of SMRT and/or TIF2 co-expressed with the retinoid CD3254 or the retinoid TTNPB, as indicated. **e-g**, Mammalian two-hybrid assays with Gal4-SMRT and RARαLBD-VP16, and the indicated ligands (**e**) or with Gal4-SMRT (**f**) Gal4-N-CoR (**g**) and RARαLBD-VP16 or RARβLBD-VP16. **h**, The retinoid CD3254 transactivates through RARβ-RXR heterodimers. Transient transactivation of DR5-tk-CAT in MCF-7 cells expressing exogenous RARα or RARβ. **i**, The retinoid CD3254 activates a retinoid target gene in H3396 cells. Transient transactivation of DR5-tk-CAT in MCF-7, H3396 or T47D cells in the presence of the ligands, indicated below. Error bars in **d-i** indicate s.e.m.

mediator for retinoic and thyroid hormone receptors (SMRT)¹⁴, which bound efficiently to the non-liganded RAR-RXR heterodimer in EMSAs and dissociated as expected in the presence of RAR agonists, remained bound when exposed to the pure rexinoid SR11237 alone (Fig. 3a). Crosslinking experiments confirmed these results and demonstrated that SMRT was able to bind to apo-RAR/RXR when SR11237 was added under conditions which also led to efficient recruitment of TIF2.42 (Fig. 3b). No RAR-RXR-SMRT-TIF2.42 complex was observed, suggesting that SMRT and TIF2 could not bind together to the heterodimer: SMRT and TIF may even compete for binding to the heterodimer. Furthermore, EMSAs carried out under similar conditions did not reveal any tetrameric complexes; increasing the concentration of SMRT caused an increase in abundance of the apo-RAR/RXR/SMRT complex, whereas with increased TIF2 concentration the apo-RAR/RXR/TIF2.42 complex became predominant (Fig. 3c and data not shown). Recent data indicating that the interaction surfaces of co-activators and co-repressors are similar¹⁵⁻¹⁷, and that a single co-repressor molecule binds to the heterodimer^{18,19}, provide the rationale for the mutually exclusive binding of co-activators and co-repressors. Indeed, altering the relative concentrations of TIF2 and SMRT in transient transfections strongly affected transactivation by the pure rexinoid CD3254 (Fig. 3d). Whether TIF2-SMRT competition in rexinoid-bound heterodimers is due to steric hindrance and/or allosteric effects remains to be established.

If the above mechanism operates in living cells, retinoid receptors interacting weakly with co-repressors should form RXR heterodimers that can activate in response to rexinoid agonists. RARβ, which when compared with RARα exhibits a largely reduced binding efficiency to nuclear receptor co-repressor (N-CoR) and SMRT (Fig. 3f, g and ref. 20), activates a DR5-based reporter gene in response to CD3254 nearly as well as TTNPB, whereas RARα is active in the presence of TTNPB but not CD3254 (Fig. 3h). Notably, a comparison between a number of breast cancer cell lines revealed that rexinoid agonists efficiently activate DR5-based reporter genes in H3396 but not MCF-7 or T47D cells (Fig. 3i). RNase mapping showed that the p160 co-activator AIB1/RAC3/ACTR is overexpressed in H3396 cells; in T47D cells expression of N-CoR is highest, whereas levels of the p160 co-activators and co-repressors are otherwise similar in both cell lines (data not shown).

Two classes of retinoid antagonists (BMS614 or AGN and BMS493) display very different modes of action. Although none of these allow efficient co-activator recruitment (Figs 1b, d, 2c and 3b), BMS614 and AGN, but not BMS493, destabilize the SMRT/apo-RAR/apo-RXR complex (Fig. 3a, b; AGN data not shown). Thus, at least two classes of RAR antagonists exist. Class I antagonists allow or even reinforce (BMS493) co-repressor binding to RAR, whereas class II antagonists (BMS614) induce co-repressor dissociation (Fig. 3e). Notably, class I antagonists will enable the heterodimer to act as a silencer, whereas class II antagonists will inhibit (but not silence) agonistic activity (Fig. 4).

The crystal structure of the RARαLBD-RXRαLBD-BMS614 complex reveals that helix H12 of RARα occupies the groove to which LXXLL NR boxes of co-activators bind^{1,21}. Amino acids of the same groove also participate in co-repressor recruitment¹⁵⁻¹⁷. We therefore assume that BMS614 and AGN direct RAR helix H12 into a position that destabilizes binding of both co-activator and co-repressors. How BMS493 impairs co-activator binding while enhancing co-repressor binding remains to be established.

Our data suggest a model of RAR-RXR heterodimer action (Fig. 4) in which each subunit can autonomously bind its agonist and generate a surface for co-activator binding. RAR agonists dissociate co-repressors and recruit co-activators, resulting in transactivation (Fig. 4b). In the presence of both RAR and RXR ligands, synergy originates from the cooperative binding of two NR

boxes of one co-activator molecule (Fig. 4c). Retinoid agonists induce co-activator recruitment to RXR, but cannot dissociate co-repressors. As the bindings of co-activators and co-repressors are mutually exclusive, a retinoid agonist-bound RXR within an apo-RAR/RXR heterodimer cannot bind a co-activator in the usual cellular environment, thus accounting for RXR subordination (Fig. 4g). Retinoids can transactivate only when the heterodimeric partner interacts weakly with co-repressors or when the co-activator expression in a particular cell dominates co-repressor content. Class II retinoid antagonists (BMS614, AGN) put RAR H12 in an unfavourable position for co-activator interaction but, depending on the ligand, H12 is not completely arrested (arrow in Fig. 4d). Retinoid recruitment of co-activators facilitates the re-positioning of H12 in the class II retinoid antagonist-bound RAR subunit and leads to transactivation (Fig. 4e), whereas class I antagonists (BMS493) stabilize co-repressor interaction and enhance silencing (Fig. 4f).

Although the p160 interaction pattern accounts fully for RXR subordination and synergy between RAR and RXR ligands, some 'permissive' heterodimers and steroid receptor homodimers show a distinct pattern of co-activator interaction. In peroxisome proliferator-activated receptor (PPAR)-RXR γ heterodimers the RXR subunit interacts autonomously in a retinoid agonist-dependent manner with p160 co-activators, whereas holo-PPAR γ recruits DRIP205 (ref. 22). The glucocorticoid²³, oestrogen²⁴ and androgen²⁵ receptor homodimers recruit p160 or DRIP150 co-activators to the amino-terminal activation function, which we did not observe for

the RAR-RXR heterodimer. Thus, distinct receptors appear to exhibit different co-regulator stoichiometries and interaction patterns that can be modulated by the use of a particular ligand²⁶ (see above). The identity and sequential establishment of various co-regulator complexes may contribute to cell- and receptor-specific gene programming. □

Methods

Materials, transfection and protein purification

Ligands have been described previously^{6,13}. BMS493 is a pan-RAR antagonist (G. Christoffel, C. Gaudon and H.G., data not shown); CD3254 is an RXR α agonist with an equilibrium dissociation constant (K_d) of 3 nM for RXR α and $>10 \mu\text{M}$ for RARs (S. Michel, personal communication). pSG5-based RAR and RXR expression vectors, reporter genes and transient transfections have been described²⁷. Gal4-SMRT corresponds to a fusion between residues 1–147 of Gal4 and residues 982–1,495 of SMRT. We used the TNT transcription/translation system (Promega) to produce proteins *in vitro*. Bacterial expression vectors were constructed in pET15b and express the following sequences: RAR α Δ AB (RAR) and RXR α Δ AB (RXR), TIF2.42 residues 624–828 (gift of L. Desaubry), TIF2.1 residues 624–1287, SMRTct (SMRT) residues 982 to end. In dnRAR α Δ AB (dnRAR) and dnRXR α Δ AB (dnRXR), the residues 408–416 and 456–467, respectively, are deleted, in addition to region AB. TIF2.42 mutants (M13, M1, M2, M3) were constructed by polymerase chain reaction (PCR)-assisted site-directed mutagenesis. Overexpression and purification were as described²⁸.

Chemical crosslinking

Purified proteins (RAR-RXR, dnRAR-RXR and RAR-dnRXR heterodimers²⁸, TIF2.42 and SMRT) were dialysed against 10 mM HEPES, pH 8.0, containing 0.1 mM EDTA, 200 mM NaCl, 1 mM dithiothreitol (DTT) and 5% glycerol. A mixture containing 5 μl RXR-RAR (5 μM) and 5 μl TIF2.42 (10 μM) was incubated with saturating amounts of ligand at 4 °C for 30 min. In a total volume of 15 μl , crosslinking reaction was carried out by addition of 0.15 μl disuccinimidyl glutarate (DSG, 10 mM) in dimethyl sulphoxide (DMSO). Mixtures were incubated for 20 min at room temperature. The crosslinking reaction was terminated by the addition of 1.5 μl Tris (1 M), pH 8.0, and the tubes were placed immediately on ice for a further 5–10 min. We added approximately 35 μl Laemmli sample buffer to the crosslinked samples. Five microlitres were resolved on 10% SDS-polyacrylamide gels followed by western blotting and detection by anti-RAR or anti-TIF2 antibodies.

Electrophoretic mobility shift assay

The heterodimers were incubated with saturating amounts of the respective ligands for 30 min at 4 °C. Co-regulators (TIF2.42, TIF2.1 or SMRT) were added and incubated for a further 15–30 min at 4 °C. The heterodimer-co-regulator complexes were further incubated for 15 min with 25,000 c.p.m. pre-annealed DR5 that was labelled with ³²P (5'-TCGAGGGTAGGGGTACCCGAAGGTCACCTCG-3'; direct repeat underlined) oligonucleotide at 4 °C. The total volume of the entire reaction mix was 22 μl . The protein-DNA complexes were resolved on non-denaturing 6% polyacrylamide gels in 0.5 $\times$ TBE buffer for 3 h (pre-run overnight at 200 V, 4 °C). The gels were dried and subjected to autoradiography.

Chromatin immunoprecipitation assays

ChIP assays were carried out with the acetyl-histone H3 immunoprecipitation assay kit (Upstate Biotechnology). Briefly, approximately 10⁶ NB4 cells were treated with ligands for 2 h and ChIP assays were done using antisera directed against acetylated histone H3 as specified by the manufacturer. Identical data were obtained when we used anti-histone H4 antibodies (data not shown). Quantitative PCR of the RAR β gene promoter was performed with immunoprecipitated chromatin using the oligonucleotide 5'-GAGAAT CCTGGGAGTTGGTGA-3' and 5'-TTGGCAAGAATAGACCCTC-3'. In Fig. 2d the relative acetylation levels are depicted as chromatin immunoprecipitated DNA that had been amplified by PCR (top panel); quantitative values for histone acetylation (bottom panel) were derived from quantitative PCR measurements of at least three independent ChIP assays.

Electrospray ionization mass spectrometry

We performed ESI-MS as described²⁹. The accelerating cone voltage (V_c) was optimized to get the best ion desolvation level for molecular mass measurements without provoking dissociation of non-covalently bound species in the gas phase. At $V_c = 120$ V molecular masses corresponding to a 1:1 ratio between the heterodimer and TIF2.42 were measured. Detection of the doubly liganded heterodimer was achieved at $V_c = 40$; under such conditions attachment of protons leads to a series of mass-to-charge (m/z) ratio peaks for a single-protein species. We were able to detect only one species of complexes corresponding to heterodimer-TIF2.42-Am80-CD3254 association but not complexes with two TIF2 molecules (not shown). The spectrum in Fig. 1h shows two series of peaks corresponding to differently charged states (15–17 protonations, series A; 20–22 protonations, series B); molecular masses of series A correspond to heterodimers, those of B to heterodimer plus a single TIF2.42. The presence of the two ligands (Am80 and CD3254) was confirmed by analyses under denaturant conditions (50% CH₃CN, 0.5% triethanolamine, TEA). These ligands correspond to molecules bound to heterodimer as the

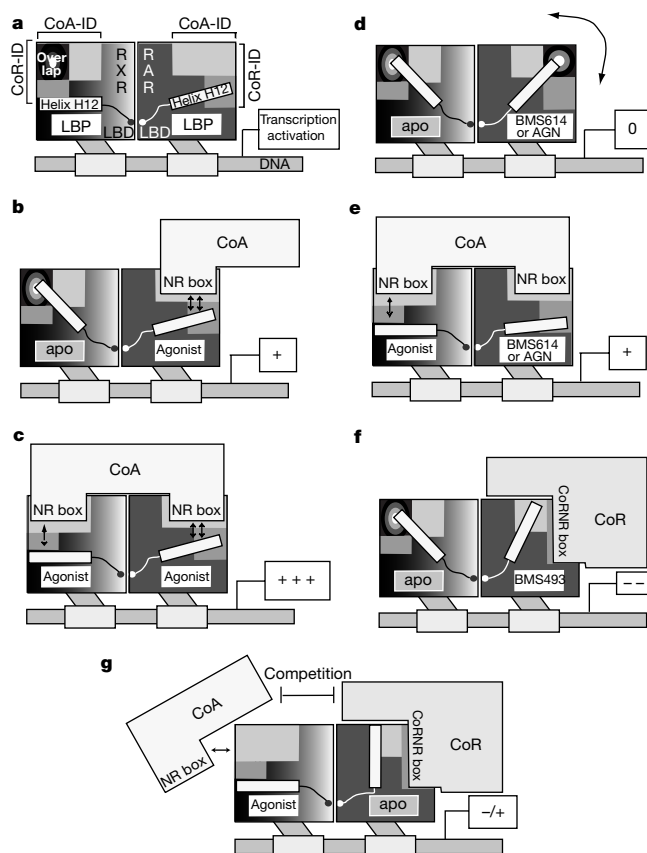


Figure 4 Proposed model for RAR-RXR heterodimer (HD) function in the presence of agonists and two different types of antagonist. **a**, Cartoon depicting the RAR-RXR heterodimer bound to a DNA response element. Within the ligand-binding domain (LBD) the ligand-binding pocket (LBP), carboxy-terminal helix H12, co-activator (CoA)- and co-repressor (CoR)-interacting domains (IDs), and their overlap, are shown. **b–g**, See text for a detailed description.

heterodimer–ligand complexes were dialysed against a buffer without ligands before ESI-MS analyses.

Received 25 July; accepted 31 October 2001.

- Bourguet, W., Germain, P. & Gronemeyer, H. Nuclear receptor ligand-binding domains: three-dimensional structures, molecular interactions and pharmacological implications. *Trends Pharmacol. Sci.* **21**, 381–388 (2000).
- Glass, C. K. & Rosenfeld, M. G. The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev.* **14**, 121–141 (2000).
- Freedman, L. P. Increasing the complexity of coactivation in nuclear receptor signaling. *Cell* **97**, 5–8 (1999).
- Kurokawa, R. *et al.* Regulation of retinoid signalling by receptor polarity and allosteric control of ligand binding. *Nature* **371**, 528–531 (1994).
- Forman, B. M., Umesono, K., Chen, J. & Evans, R. M. Unique response pathways are established by allosteric interactions among nuclear hormone receptors. *Cell* **81**, 541–550 (1995).
- Chen, J. Y. *et al.* Two distinct actions of retinoid-receptor ligands. *Nature* **382**, 819–822 (1996).
- Vivat, V. *et al.* A mutation mimicking ligand-induced conformational change yields a constitutive RXR that senses allosteric effects in heterodimers. *EMBO J.* **16**, 5697–5709 (1997).
- Roy, B., Taneja, R. & Chambon, P. Synergistic activation of retinoic acid (RA)-responsive genes and induction of embryonal carcinoma cell differentiation by an RA receptor alpha (RAR α)-, RAR β -, or RAR γ -selective ligand in combination with a retinoid X receptor-specific ligand. *Mol. Cell. Biol.* **15**, 6481–6487 (1995).
- Voegel, J. J. *et al.* The coactivator TIF2 contains three nuclear receptor-binding motifs and mediates transactivation through CBP binding-dependent and -independent pathways. *EMBO J.* **17**, 507–519 (1998).
- Lehmann, J. M. *et al.* Retinoids selective for retinoid X receptor response pathways. *Science* **258**, 1944–1946 (1992).
- Voegel, J. J., Heine, M. J. S., Zechel, C., Chambon, P. & Gronemeyer, H. TIF2, a 160 kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors. *EMBO J.* **15**, 3667–3675 (1996).
- Westin, S. *et al.* Interactions controlling the assembly of nuclear-receptor heterodimers and co-activators. *Nature* **395**, 199–202 (1998).
- Klein, E. S. *et al.* Identification and functional separation of retinoic acid receptor neutral antagonists and inverse agonists. *J. Biol. Chem.* **271**, 22692–22696 (1996).
- Chen, J. D. & Evans, R. M. A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature* **377**, 454–457 (1995).
- Hu, X. & Lazar, M. A. The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors. *Nature* **402**, 93–96 (1999).
- Nagy, L. *et al.* Mechanism of corepressor binding and release from nuclear hormone receptors. *Genes Dev.* **13**, 3209–3216 (1999).
- Perissi, V. *et al.* Molecular determinants of nuclear receptor-corepressor interaction. *Genes Dev.* **13**, 3198–3208 (1999).
- Hu, X., Li, Y. & Lazar, M. A. Determinants of CoRNR-dependent repression complex assembly on nuclear hormone receptors. *Mol. Cell. Biol.* **21**, 1747–1758 (2001).
- Cohen, R. N. *et al.* The specificity of interactions between nuclear hormone receptors and corepressors is mediated by distinct amino acid sequences within the interacting domains. *Mol. Endocrinol.* **15**, 1049–1061 (2001).
- Wong, C. W. & Privalsky, M. L. Transcriptional silencing is defined by isoform- and heterodimer-specific interactions between nuclear hormone receptors and corepressors. *Mol. Cell. Biol.* **18**, 5724–5733 (1998).
- Bourguet, W. *et al.* Crystal structure of a heterodimeric complex of RAR and RXR ligand-binding domains. *Mol. Cell* **5**, 289–298 (2000).
- Yang, W., Rachez, C. & Freedman, L. P. Discrete roles for peroxisome proliferator-activated receptor gamma and retinoid X receptor in recruiting nuclear receptor coactivators. *Mol. Cell. Biol.* **20**, 8008–8017 (2000).
- Hittelman, A. B., Burakov, D., Iniguez-Lluhi, J. A., Freedman, L. P. & Garabedian, M. J. Differential regulation of glucocorticoid receptor transcriptional activation via AF-1-associated proteins. *EMBO J.* **18**, 5380–5388 (1999).
- Benecque, A., Chambon, P. & Gronemeyer, H. Synergy between estrogen receptor α activation functions AF1 and AF2 mediated by transcription intermediary factor TIF2. *EMBO Rep.* **1**, 151–157 (2000).
- Alen, P., Claessens, F., Verhoeven, G., Rombauts, W. & Peeters, B. The androgen receptor amino-terminal domain plays a key role in p160 coactivator-stimulated gene transcription. *Mol. Cell. Biol.* **19**, 6085–6097 (1999).
- Yang, W. & Freedman, L. P. 20-Epi analogues of 1,25-dihydroxyvitamin D3 are highly potent inducers of DRIP coactivator complex binding to the vitamin D3 receptor. *J. Biol. Chem.* **274**, 16838–16845 (1999).
- Nagpal, S., Friant, S., Nakshatri, H. & Chambon, P. RARs and RXRs: evidence for two autonomous transactivation functions (AF-1 and AF-2) and heterodimerization *in vivo*. *EMBO J.* **12**, 2349–2360 (1993).
- Iyer, J. *et al.* Versatile copurification procedure for rapid isolation of homogeneous RAR-RXR heterodimers. *Protein Expr. Purif.* **16**, 308–314 (1999).
- Bourguet, W. *et al.* Heterodimeric complex of RAR and RXR nuclear receptor ligand-binding domains: purification, crystallization, and preliminary X-ray diffraction analysis. *Protein Expr. Purif.* **19**, 284–288 (2000).
- Chen, J. *et al.* RAR-specific agonist antagonists which dissociate transactivation and AP1 transrepression inhibit anchorage-independent cell proliferation. *EMBO J.* **14**, 1187–1197 (1995).

Acknowledgements

We thank W. Raffelsberger for providing the PCR protocols for RAR β 2 promoter amplification, and N. Potier and A. van Dorsselaar for the ESI-MS analysis. We are grateful to C. Peluso-Iltis and V. Chavant for help in protein purification, C. Erb for expression vectors and A. Pornon for technical assistance. We thank W. Bourguet for critically reading this manuscript and Y. Lutz and C. Egly for antibodies. BMS-labelled retinoids/rexinoids

were provided by C. Zusi, and CD3254 and AGN192870 by U. Reichert and S. Michel. This work was supported by the Institut National de la Santé et de la Recherche Médicale, the Centre National de La Recherche Scientifique, the Hôpital Universitaire de Strasbourg, the European Community and Bristol-Myers Squibb.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.G. (e-mail: hg@igbmc.u-strasbg.fr).

The motor domain determines the large step of myosin-V

Hiroyo Tanaka*, Kazuaki Homma†, Atsuko Hikikoshi Iwane‡, Eisaku Katayama§||, Reiko Ikebe†, Junya Saito†, Toshio Yanagida*‡ & Mitsuo Ikebe†

* Single Molecule Processes Project, ICORP, JST, 2-4-14, Senba-higashi, Mino, Osaka 562-0035, Japan

† Department of Physiology, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, Massachusetts 01655-0127, USA

‡ Department of Physiology and Biosignaling, Graduate School of Medicine Osaka University, 2-2, Yamadaoka, Suita, Osaka, 565-0871, Japan

§ Division of Biomolecular Imaging, Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, 108-8639, Japan

|| PRESTO, Japan Science and Technology Corporation, Kawaguchi, Saitama 332-0012, Japan

Class-V myosin proceeds along actin filaments with large (~36 nm) steps^{1–3}. Myosin-V has two heads, each of which consists of a motor domain and a long (23 nm) neck domain. In accordance with the widely accepted lever-arm model⁴, it was suggested that myosin-V steps to successive (36 nm) target zones along the actin helical repeat by tilting its long neck (lever-arm)⁵. To test this hypothesis, we measured the mechanical properties of single molecules of myosin-V truncation mutants with neck domains only one-sixth of the native length. Our results show that the processivity and step distance along actin are both similar to those of full-length myosin-V. Thus, the long neck domain is not essential for either the large steps or processivity of myosin-V. These results challenge the lever-arm model. We propose that the motor domain and/or the actomyosin interface enable myosin-V to produce large processive steps during translocation along actin.

Structural studies of myosin in crystal⁶ and solution^{7,8} have shown that the angle of the neck domain of the myosin head relative to the motor domain changes, depending on the form of bound nucleotide, as a result of the conformational change in the motor domain. On the basis of these findings, a lever-arm model has been proposed, in which the neck domain acts as a lever-arm and the movement is caused by the tilting neck domain⁴. This model postulates that the displacement and the velocity of myosin are proportional to the length of the neck domain. To test this model, the displacement and sliding velocity of intact and biochemically modified myosin, which have different lengths of neck domains, has been measured. However, not all of the results are consistent with this model; some investigators have shown that the displacement^{9,10} and velocity^{9,11} are proportional to the neck domain length, but others have shown that the neck domain length does not affect the velocity^{12–15}. Myosin-V, which has a long neck domain (two to three times as large as those of conventional myosin-II), moves on an actin filament for a long distance with many successive large steps (~36 nm) without dissociating from actin, called processive

movement^{1,3}. The large steps and processivity of myosin-V make it possible to determine the displacement and movement of individual myosin-V molecules more directly and precisely. We thus studied the mechanism of myosin movement using myosin-V and its truncation mutants.

We constructed a chimaera of the head region of myosin-V attached to the chicken smooth-muscle myosin rod, in which five out of six calmodulin-binding 'IQ-domains' (IQ-motifs) in the neck region were truncated (M5IQ1rod; Fig. 1a). The actin translocating velocity, which is driven by multiple M5IQ1rods bound to a glass surface (surface assay), was $0.30 \pm 0.10 \mu\text{m s}^{-1}$ ($\pm$ standard deviation; $n = 132$). This was similar to previously reported actin velocities driven by wild-type myosin-V (ref. 1) and recombinant myosin-V fragments—HMM and S1—fully-equipped with six IQ-motifs^{13,15,16}. For control experiments, a M5IQ6rod containing six intact IQ-motifs, a long neck of myosin-V, and a smooth-muscle myosin rod was also constructed (Fig. 1b). M5IQrods and M5IQ6rods were co-polymerized into 5–8- μm filaments with rabbit skeletal-muscle myosin rods. During the co-polymerization,

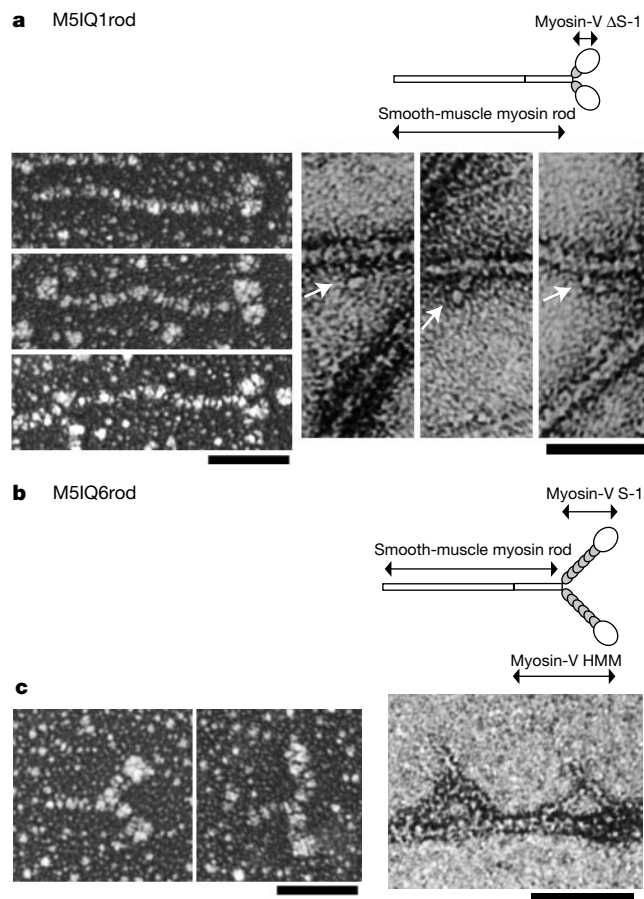


Figure 1 Structure of the deletion mutant of myosin-V (M5IQ1rod). **a**, Diagram of the M5IQ1rod (top). The M5IQ1rod consists of the two deletion heads of myosin-V (a motor domain plus one IQ-motif for each head (Δ S-1)) and the long coiled-coil tail (rod) of chicken smooth-muscle myosin. Rotary-shadowed electron microscopic images of M5IQ1rods (lower left). Negatively stained images of M5IQ1rods interacting with actin filaments (lower right). Myosins interacted with actin filaments in assay buffer containing approximately $5 \mu\text{M}$ ATP, then the assay buffer was rapidly replaced by staining solution²⁸. The arrows indicate M5IQrod molecules interacting with actin filaments. Scale bars indicate 50 nm. **b**, Diagram of the M5IQ6rod (M5IQ6 contains six intact IQ-motifs). **c**, Rotary-shadowed electron microscopic images of myosin-V HMM (see **b**) (left) and negatively stained images of its complexes with actin filaments (right). The negatively stained images, taken under the same conditions as those in **a**, were consistent with those of myosin-V HMM previously reported⁹.

we incubated the skeletal-muscle myosin rods at larger molar excess than for the M5IQ1rods or M5IQ6rods to ensure that only a few M5IQ1rods or M5IQ6rods were included in each coflament. Cofilaments were sparsely adsorbed onto the surface of a pedestal made on a glass surface^{17,18} (Fig. 2a). The number of M5IQ1rods or M5IQ6rods in a coflament was checked by observing single fluorescent nucleotide analogues (Cy3-ATP or Cy3-ADP) bound to M5IQ1rods or M5IQ6rods^{17,18} (Fig. 2b). The use of long cofilaments not only allowed us easily to find the location of M5IQ1rods or M5IQ6rods on the glass surface, but also avoided any damage to them caused by interaction with the glass. An actin filament with both ends attached to optically trapped beads was brought into contact with single M5IQ1rods or M5IQ6rods in a coflament, and individual displacement events were determined by measuring bead displacements with nanometre accuracy (Fig. 2a) as previously described^{17–20}.

We observed the mechanical events of single M5IQ1rods and M5IQ6rods at 2 mM ATP. Figure 3a and b show typical recordings of the displacements of M5IQ1rods and M5IQ6rods, respectively. Displacements occurred in steps (Fig. 3a, b, arrows). Most of the displacements of both myosins took place in one direction (defined as forward). In the 'falling' phase, many ($\sim 60\%$) of the displacements abruptly returned to zero displacement without steps (Fig. 3a left, b), indicating that the myosin dissociated from the actin filament, while others returned to zero in a stepwise fashion (Fig. 3a, right). The start position of the displacement cannot be determined because it is randomized owing to the thermal motion of the beads²⁰. Therefore, the size of the first step was determined by averaging many steps, in which the start position was averaged to be zero displacement²⁰ (insets to Fig. 3a left and b). The averaged size of the first step was 25 nm for a M5IQ1rod and 20 nm for a M5IQ6rod. The start positions of the second and third steps could be determined (Fig. 3a, b), so the size of individual steps was measured directly. Figure 3c shows a histogram of the step size of a M5IQ1rod in the forward and backward directions (the first steps were not included). The mean sizes of the forward and backward steps were both approximately 35 nm. Most of the backward steps took place in

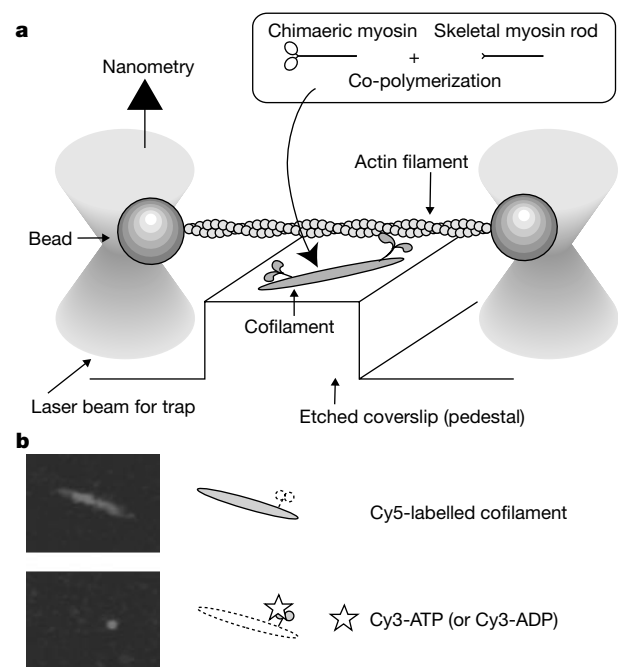


Figure 2 Single-molecule measurement of displacements. **a**, Diagram of the measurement system for displacements. **b**, Fluorescence images (with corresponding diagrams) of a coflament (top) and Cy3 nucleotides bound to M5IQ1rod in the coflament (bottom); see text. A single M5IQ1rod molecule was incorporated into a coflament.

the falling phase, that is, passively (Fig. 3c, grey bars), but a small number of backward steps developed against the force owing to the optically trapped bead (Fig. 3c, black bars). The ratio of forward steps to backward steps that developed against the force (Fig. 3c, black bars) was 10:1. We observed 56 displacements in which two successive 35-nm steps took place against the trap force in the forward direction, but we never observed such a displacement in the backward direction.

Figure 3d shows a histogram of dwell times between the adjacent forward steps of a M5IQ1rod. The mean dwell time for forward steps was 0.11 s. The velocity of forward displacements calculated as (mean step size)/(mean dwell time) was equal to 320 nm s^{-1} ($= 35 \text{ nm}/0.11 \text{ s}$). This value was consistent with that in the surface-gliding assay for a M5IQ1rod. The dwell times for the last steps were not included because the optical trap force stalled the movement of the M5IQ1rod (Fig. 3a, b).

The step size, processivity, and the dwell time of a M5IQ6rod were all similar to those of a M5IQ1rod (Fig. 3b). The step sizes (35 nm) of M5IQ1rods and M5IQ6rods were consistent with that of wild-type myosin-V (refs 1, 3). The result clearly indicates that the large lever-arm length of myosin-V is not a determinant of its large step size. The M5IQ6rod underwent backward steps similarly to the M5IQ1rod. Backward steps have been reported for wild-type myosin-V (refs 1, 3) as well.

A maximum number of three processive steps was observed; back steps were sometimes observed. Therefore, it could be argued that the large noise in Fig. 3a and b is caused by the thermal forwards and backwards bouncing motion of the actin filament when it is released from myosin, so the myosin just binds to appropriate monomers presented in the correct orientation every 36 nm. This might be the case, if the load is nearly zero, but the trapping force exerted on the myosin was large in our assay. Because the trap stiffness was 0.02 pN nm^{-1} , if the starting position is zero displacement, the

force increases from 0 to 0.4 pN at the first step, from 0.4 to 1.1 pN at the second step and from 1.1 to 1.8 pN at the third step. The energy required for generating the fourth step, if possible, is given as $35 \text{ nm} \times (1.8 + 2.5) \text{ pN}/2 = 75 \text{ pN nm}^{-1}$, which is similar to the energy of ATP hydrolysis²¹. If the efficiency is similar to that of muscle myosin ($\sim 50\%$), it is reasonable not to expect more than three processive steps.

Furthermore, the probability that two successive 35-nm steps are thermally produced against the force of over 1 pN in one direction is extremely small: less than $\exp(-35 \text{ nm per step} \times 2 \text{ steps} \times 1 \text{ pN}/k_B T) = 10^{-8}$. We observed two such successive 35-nm steps 56 times, but only in the forward direction. Thus, the observed steps must be actively caused using the energy from ATP hydrolysis. The noise in Fig. 3a and b can be explained by the large compliance ($1/\text{stiffness}$) of the bead-actin filament junction²². That is why the noise did not decrease much when the myosin attached to an actin filament.

Many (more than 3) processive steps of wild-type myosin-V have been observed³. This would not be because the wild-type myosin-V was used, but because the load was low (the load was controlled to be relatively small (1 pN) using feedback-enhanced optical trapping nanometry). A wild-type myosin-V also produced a number of steps similar to the present study when the steps were measured by optical trapping nanometry without feedback¹.

When the concentration of ATP was decreased to $[\text{ATP}] = 1 \mu\text{M}$, the displacements of M5IQ1rods also developed in a stepwise fashion as well (data not shown). The size of step (mean, 35 nm) was similar to that at 2 mM . The dwell time, however, increased to 1.7 s . At $1 \mu\text{M}$ ATP, the binding of ATP to an acto-M5IQ1rod complex would be a rate-limiting step. The second-order binding constant of ATP calculated as $1/[\text{ATP}](\text{dwell time}) = 0.6 \mu\text{M}^{-1} \text{ s}^{-1}$. This value is consistent with the kinetic analyses for myosin-V with six IQ motifs^{3,13,15,16,23}.

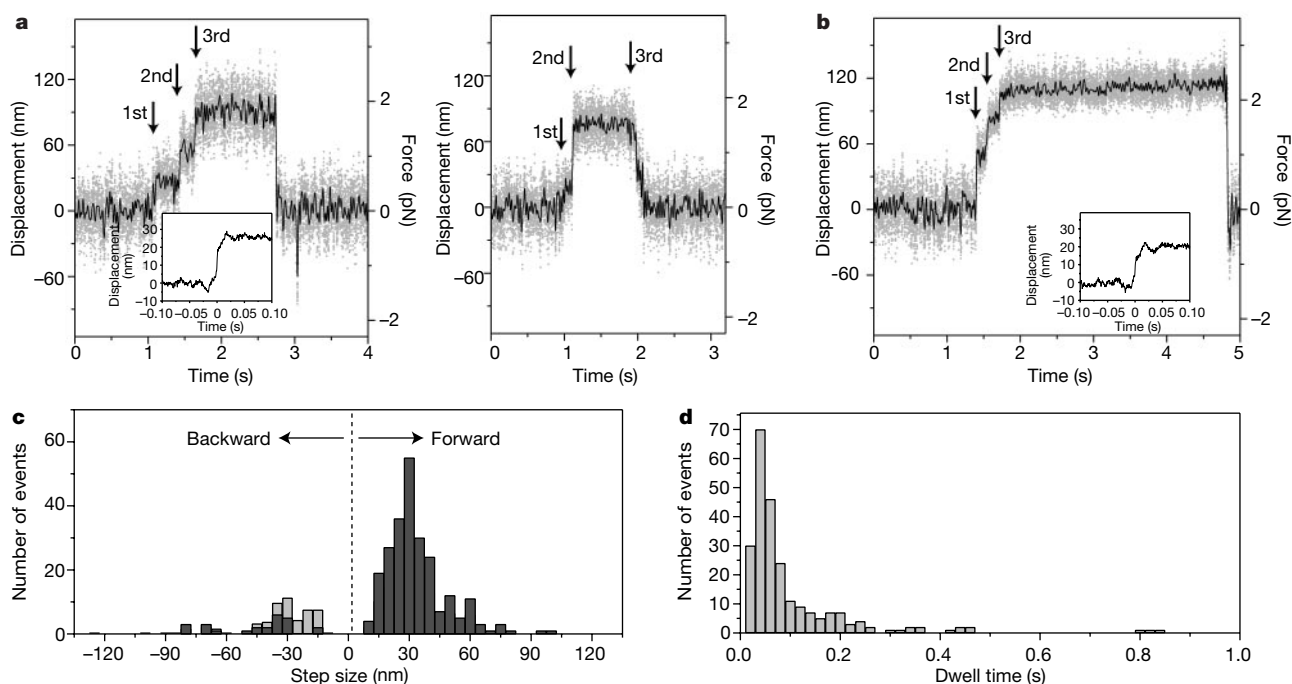


Figure 3 Records of displacements at a high ATP concentration of 2 mM . Typical records of the time courses of generation of displacements by single M5IQ1rods (**a**, right and left) and M5IQ6rods (**b**). Grey dots, raw data. Black line, data passed through a low-pass filter of 30-Hz bandwidth. Arrows indicate positions of steps. Insets in **a** and **b** are averaged traces of first steps (see text). **c**, Histogram of step size of a M5IQ1rod. In the backward direction, black and grey bars indicate the numbers of steps that developed against the force owing to the optically trapped bead and in the falling phase (passive steps),

respectively (see text for detail). The large step sizes of $>60 \text{ nm}$ are probably due to two successive steps that took place rapidly within the temporal resolution (1.6 ms) of the measurement system (see the second step in **a**, right). **d**, Histogram of dwell times of steps of M5IQ1rod in the forward direction. Medium: 25 mM KCl , 5 mM MgCl_2 , 1 mM EGTA , 0.2 mg ml^{-1} calmodulin, 2 mM ATP and 20 mM HEPES ($\text{pH } 7.8$). The medium contained $4 \mu\text{M}$ phalloidin to stabilize actin filaments and an oxygen scavenger system to reduce photobleaching¹⁷. Trap stiffness, 0.02 pN nm^{-1} . Temperature, 30°C .

Thus the size, processivity, and velocity of steps of the M5IQ1rods were essentially the same as those obtained for M5IQ6rods and previously reported for intact myosin-V (ref. 1) fully equipped with six IQ motifs. The step size of myosin-V coincides with the actin helical pitch (36 nm) and its two heads span the actin helical pitch of actin filaments in electron micrographs⁵. We also observed this in images of myosin-V equipped with an entire IQ domain (see Fig. 1c right). On the basis of these findings, a model was proposed for the processive movement of myosin-V in which myosin-V steps along the helical repeat of an actin filament by tilting the long neck domain of one head and leading the partner head to the neighbouring helical pitch⁵. This model is consistent with the lever-arm model⁴. However, the M5IQ1rod cannot step processively in this way, because its neck domain is too short to span the actin helical pitch. In fact, electron microscopy revealed that most M5IQ1rod molecules appeared to bind to an actin filament with either head and the heads of single molecules do not span the helical pitch (Fig. 1a right).

How can the M5IQ1rod, with its short neck domain, develop such large and processive steps? Perhaps an M5IQ1rod kicks the actin filament to bring the next binding site to it. This is, however, impossible because the inertia of the actin filament and beads is negligibly small. The two heads of the M5IQ1rod are connected at the far ends of their neck domains, away from the actin-binding site, to a coiled-coil rod. Therefore, it is possible that if the coiled coil is untwisted, the two heads could span the helical pitch of an actin filament. However, this is unlikely to occur, because electron microscopy revealed that the two heads of an M5IQ1rod, both in the presence and absence of actin filaments, were rigidly connected at their necks (Fig. 1a, lower left and lower right, respectively). Thus, the M5IQ1 rod does not appear to step along the helical repeats of an actin filament.

The most probable mechanism for the movement of the M5IQ1rod is that the head continuously and rapidly slides along an actin filament to develop an overall large step. The rotation of actin within a step (10 ms) is less than 15 degrees, using the rotational stiffness of an actin filament and the rotational drag of beads²⁴. Therefore, the head of an M5IQ1rod fixed on a glass surface could not interact continuously with one protofilament of a double-stranded helix of an actin filament. Consequently, to move continuously between the helical repeats, the M5IQ1rod molecule needs to change the tracks from one protofilament to the other, at the cross-point of the helix. It is plausible that M5IQ1rod, having two heads, slides on the one actin protofilament with one of the two heads, and switches over to the other protofilament with the partner head at the cross-point. The fact that the step size coincides with the helical repeat (36 nm) may be because the sliding of one head of M5IQ1rod on an actin protofilament pauses at the cross-point. The unidirectional movement of myosin heads along the actin helical pitch may be caused through a potential slope made along the helical pitch of an actin filament via its conformational change upon myosin binding. □

Methods

Proteins

Actin and myosin rods were obtained from rabbit skeletal muscle and purified¹⁷. Recombinant calmodulin from *Xenopus* oocytes was expressed in *Escherichia coli* as described²⁵.

Generation of the expression vectors for myosin-V constructs

Mouse myosin-V complementary DNA clones were supplied by N. Jenkins²⁶. A baculovirus transfer vector for mouse myosin-V variants in pBluebac4 (Invitrogen) was produced as follows. A cDNA fragment (1–3,578) was excised with *NheI/KpnI* digestion and in-frame ligated to a pBluebac4His baculovirus transfer vector containing a hexahistidine tag sequence with a stop codon at the 3' side of the *KpnI* site (M5IQ6 HMM). To obtain single headed myosin-V with one IQ-motif, a *KpnI* site was created at nucleotide 2,392. The vector was digested with *KpnI* then self-ligated (M5IQ1S1). To produce M5IQ1rod, a *SpeI* site was created at the 5' side of the unique *KpnI* site of M5IQ1S1 and

chicken smooth muscle myosin heavy-chain cDNA fragment (3,325–5,812) flanked with a *SpeI* and a *KpnI* site was inserted to *SpeI/KpnI* sites at the 3' side of M5IQ1S1. To produce M5IQ6rod, a *SpeI* site was created at nucleotide 2,770 of M5IQ6 HMM and the cDNA fragment encoding a smooth-muscle myosin rod flanked with a *SpeI* and a *KpnI* site was inserted into the *SpeI/KpnI* site of M5IQ6 HMM. For all constructs, a *SpeI* site was mutated to restore the authentic amino acid residues.

Preparation of recombinant myosin-V

to express recombinant myosin-V, Sf9 cells (about 1×10^9) were co-infected with two separate viruses expressing the myosin-V heavy chain and calmodulin, respectively, as previously described²⁷. No contamination of endogenous myosin-V was observed in an SDS–polyacrylamide gel electrophoresis (PAGE) or in rotary-shadowed electron microscopy observation of over 200 molecules.

Received 1 August; accepted 18 October 2001.

- Mehta, A. D. *et al.* Myosin-V is a processive actin-based motor. *Nature* **400**, 590–593 (1999).
- Sakamoto, T., Amitani, L., Yokota, E. & Ando, T. Direct observation of processive movement by individual myosin V molecules. *Biochem. Biophys. Res. Commun.* **272**, 586–590 (2000).
- Rief, M. *et al.* Myosin-V stepping kinetics: A molecular model for processivity. *Proc. Natl Acad. Sci. USA* **97**, 9482–9486 (2000).
- Spudich, J. A. How molecular motors work. *Nature* **372**, 515–518 (1994).
- Walker, M. L. *et al.* Two-headed binding of a processive myosin to F-actin. *Nature* **405**, 804–807 (2000).
- Cooke, R. Myosin structure: does the tail wag the dog? *Curr. Biol.* **9**, R773–R775 (1993).
- Wakabayashi, K. *et al.* Small-angle synchrotron X-ray scattering reveals distinct shape changes of the myosin head during hydrolysis of ATP. *Science* **258**, 443–447 (1992).
- Corrie, J. E. *et al.* Dynamic measurement of myosin light-chain-domain tilt and twist in muscle contraction. *Nature* **400**, 425–430 (1999).
- Warshaw, D. M. *et al.* The light chain binding domain of expressed smooth muscle heavy meromyosin acts as a mechanical lever. *J. Biol. Chem.* **275**, 37167–37172 (2000).
- Ruff, C., Furch, M., Brenner, B., Manstein, D. J. & Meyhofer, E. Single-molecule tracking of myosins with genetically engineered amplifier domains. *Nature Struct. Biol.* **8**, 226–229 (2001).
- Uyeda, T. Q., Abramson, P. D. & Spudich, J. A. The neck region of the myosin motor domain acts as a lever arm to generate movement. *Proc. Natl Acad. Sci. USA* **93**, 4459–4464 (1996).
- Itakura, S. *et al.* Force-generating domain of myosin motor. *Biochem. Biophys. Res. Commun.* **196**, 1504–1510 (1993).
- Trybus, K. M., Kremontsova, E. & Frey, Y. Kinetic characterization of a monomeric unconventional myosin V construct. *J. Biol. Chem.* **274**, 27448–27456 (1999).
- Perreault-Micale, C., Shushan, A. D. & Coluccio, L. M. Truncation of a mammalian myosin I results in loss of Ca^{2+} -sensitive motility. *J. Biol. Chem.* **275**, 21618–21623 (1999).
- Homma, K., Saito, J., Ikebe, R. & Ikebe, M. Ca^{2+} -dependent regulation of the motor activity of myosin V. *J. Biol. Chem.* **275**, 34766–34771 (2000).
- Wang, F. *et al.* Effect of ADP and ionic strength on the kinetic and motile properties of recombinant mouse myosin V. *J. Biol. Chem.* **275**, 4329–4335 (2000).
- Tanaka, H., Ishijima, A., Honda, M., Saito, K. & Yanagida, T. Orientation dependence of displacements by a single one-headed myosin relative to the actin filament. *Biophys. J.* **75**, 1886–1894 (1998).
- Ishijima, A. *et al.* Simultaneous observation of individual ATPase and mechanical events by a single myosin molecule during interaction with actin. *Cell* **92**, 161–171 (1998).
- Finer, J. T., Simmons, R. M. & Spudich, J. A. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature* **368**, 113–119 (1994).
- Molloy, J. E., Burns, J. E., Kendrick-Jones, J., Tregear, R. T. & White, D. C. Movement and force produced by a single myosin head. *Nature* **378**, 209–212 (1995).
- Stryer, L. *Biochemistry* 4th edn, 443–462 (Freeman, New York, 1995).
- Dupuis, D. E., Guilford, W. H., Wu, J. & Warshaw, D. M. Actin filament mechanics in the laser trap. *J. Muscle Res. Cell Motil.* **18**, 17–30 (1997).
- De La Cruz, E. M., Wells, A. L., Rosenfeld, S. S., Ostap, E. M. & Sweeney, H. L. The kinetic mechanism of myosin V. *Proc. Natl Acad. Sci. USA* **96**, 13726–13731 (1999).
- Tsuda, Y., Yasutake, H., Ishijima, A. & Yanagida, T. Torsional rigidity of single actin filaments and actin-actin bond breaking force under torsion measured directly by in vitro micromanipulation. *Proc. Natl Acad. Sci. USA* **93**, 12937–12942 (1996).
- Ikebe, M. *et al.* A hinge at the central helix of the regulatory light chain of myosin is critical for phosphorylation-dependent regulation of smooth muscle myosin motor activity. *J. Biol. Chem.* **273**, 17702–17707 (1998).
- Mercer, J. A., Seperack, P. K., Strobel, M. C., Copeland, N. G. & Jenkins, N. A. Novel myosin heavy chain encoded by murine dilute coat colour locus. *Nature* **349**, 709–713 (1991).
- Homma, K., Yoshimura, M., Saito, J., Ikebe, R. & Ikebe, M. The core of the motor domain, not the lever-arm/converter domain, determines the direction of myosin movement. *Nature* **412**, 831–834 (2001).
- Katayama, E. The effects of various nucleotides on the structure of actin-attached myosin subfragment-1 as studied by quick-freeze deep-etch electron microscopy. *J. Biochem.* **106**, 751–770 (1989).

Acknowledgements

We are grateful to N. Jenkins for providing cDNA fragments of mouse myosin-V. We thank colleagues at the Single Molecule Processes Project, University of Massachusetts Medical School, and Osaka University for discussions, and J. West for critically reading the manuscript. This work was supported by grants from the National Institute of Health to M.I.

Correspondence and requests for materials should be addressed to T.Y. (e-mail: yanagida@phys1.med.osaka-u.ac.jp) or M.I. (e-mail: mitsuo.ikebe@umassmed.edu).

January sales

A miscellany from recent product launches.



Millistak+

Millipore www.millipore.com

Gain a little advantage

Millistak+ self-contained, disposable capsule filters have a surface area of only 23 cm² and can be used by process developers and small-scale manufacturers to clarify and pre-filter fluids. They incorporate the same media used in Millistak+ stacked cartridges, making them suitable for scale-up applications. Filter media types are available for a variety of biopharmaceutical requirements, including removal of cell debris, colloidal contaminants and trace impurities that cause haze, discolouration and odour.

ArrayPlate

High Throughput Genomics

www.htgenomics.com

Spatial awareness

The ArrayPlate multiplexed molecular profiling system allows researchers to measure many genes and proteins simultaneously using spatial arrays in microplates. The sample throughput of 96-well and 384-well microplates is used with array technology, combined with nuclease protection to provide DNA and RNA multiplexed profiling. Sensitivity and reproducibility are said to be equal to those achieved with conventional biochemical assays, at 1,000 cells on 0.01 µg total RNA and 3–13% C.V. respectively. No separations are required, and the reagents can be used to process samples for simultaneous measurement of DNA, RNA and protein using a universal assay platform. Measurements are made by imaging the entire plate to provide a read-out of all the data points.

C1 Digital Eclipse

Nikon www.nikonusa.com

Confocal microscope system

This new microscope system from Nikon features a full array of confocal imaging techniques in multiple channels. The scanning



Go Digital: Nikon's new C1 confocal microscope.

head fits a variety of microscopes and the detection and laser modules are pre-calibrated. The C1 system supports simultaneous three-channel confocal fluorescence, confocal fluorescence plus diascopic DIC, time-lapse recording and spatial analysis. Scanning capabilities include 2D (XY< XZ< and XT), 3D (XYZ<XYT) and 4D (XYZT), as well as moveable region-of-interest and bidirectional scanning. The device has a tube lens focal length of 200 mm, an objective with a parfocal distance of 60 mm and a 25 mm objective thread to achieve higher numerical apertures and longer working distances.

UltraClean

Mo Bio www.mobio.com

Eliminate endotoxins

This kit can be used to isolate supercoiled plasmid DNA from cultures of recombinant *E. coli* with <0.1 Eu per mg endotoxin levels. Endotoxins are removed with a precipitating agent that will not affect plasmid yield or purity. No toxic residues remain in the final preparation.

Gene Expression Analysis

Biomax Informatics AG www.biomax.de

Suite dreams

The Biomax Gene Expression Analysis Suite is designed to perform comprehensive analysis of complex data sets using available gene annotation. Users can quickly find groups of tightly co-expressed genes, evaluate active metabolic pathways and predict interactions among the corresponding proteins. The software reduces complexity of data by clustering genes based on their experimental expression profiles, displays detailed expression information for any select-

ed cluster and provides links to annotations for each gene using the hierarchical Biomax functional catalogue to characterize protein function. Genes of interest are placed in a biochemical context by constructing both pathway models and protein interaction networks.

PCR Cloning Kits

Qiagen www.qiagen.com

A quicker way of doing it

Qiagen's PCR Cloning Kits use UA hybridization and Qiagen EZ Competent cells to clone PCR products in 40 minutes. The PCR product is mixed directly with the pDrive Cloning Vector and the ready-to-use ligation master mix, incubated for 30 minutes, and then added to the cells for transformation. Transformed cells can be plated directly onto agar/ampicillin plates without a recovery incubation. The procedure is said to be much faster than TA-based, topoisomerase-mediated and conventional sticky- and blunt-end cloning methods, and to provide higher cloning efficiency than TA-based procedures. The pDrive Cloning Vector contains 3' U, which has a higher specificity for the A overhang of PCR products generated by *Taq* and other non-proofreading DNA polymerases.

HST-Rotor

Retrotech www.retrotech-online.de

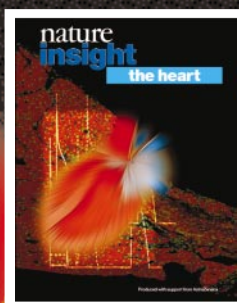
High speed, high throughput

The HST-Rotor system from Retrotech consists of a fixed-angle rotor, adapter cassettes, tubes and sealing accessories. It can be used for sample preparation of small biological specimens such as viruses, cellular organelles, nucleic acids, proteins and macromolecular complexes by centrifugation of up to 192 samples at up to approximately 40,000g. Sample volumes of 300 to 2,000 ml can be transferred rapidly from single tubes, microplates or deep well plates in a microplate eight-strip format to special eight-strip centrifugation tubes loaded into rotor cassettes in a microplate 2 × 8 strip format. The rotor fits the Sigma 3K30 and Beckmann Avanti T30 high-speed tabletop centrifuges. Depending on application, sample preparation and centrifugation can be completed within 30 minutes to two hours. Applications include transfusion medicine (HCV, HBV, HIV-nucleic acid preparation and diagnostics), HIV viral load quantification, virus research and other areas requiring high-speed centrifugation and high sample throughput.

These notes are compiled in the Nature office from information provided by the manufacturers.

nature insight

the heart



Cover illustration
Behind the cartoon of the heart is a diagram of electrical activity during atrial arrhythmic generation (courtesy S. Nattel) and stained myocardium (courtesy P. Anversa).

Ever since William Harvey defined an animal's heart as "the sovereign of everything within them, the sum of their microcosm," in *De Motu Cordis* in 1628, the heart has been one of the most widely studied organs of the body — and fortunately so, given that heart disease is the world's leading cause of death.

Each year, around 8 million people die from heart attacks and many millions more suffer from, and eventually succumb to, heart diseases such as congestive heart failure and arrhythmia. Global figures are rising, yet calculations suggest that this number could be slashed by around 50 per cent if smoking were removed from the equation. However, decreases would be offset to some degree as the world's population ages, diets become more fat-laden and lifestyles more sedentary, as all these factors are detrimental to a healthy heart.

Although scientific study of the heart began four centuries ago, the past few decades have seen a paradigm shift in research. We are now able to monitor the process of contraction and relaxation that underlies the gross function of the heart at close quarters by tracking the movement of calcium and other ions within myocytes.

Ultrastructural examination of ion channels is revealing how the regular rhythmic beating of the heart can become disordered. And genetic information is not only uncovering the basis of conditions such as cardiomyopathy — in which the heart becomes weak and unable to fulfil its blood circulatory role — but is also being exploited with relative success in trials of cardiac gene therapy. Thus, as with all diseases, understanding the processes involved at the molecular and genetic level is enabling us to make inroads in preventing and treating heart disease.

We are pleased to acknowledge the financial support of AstraZeneca in producing this Insight. As always, *Nature* carries sole responsibility for all editorial content and peer review.

Karen Birmingham

review articles:

198 Cardiac excitation-contraction coupling
D. M. Bers

206 Seven-transmembrane-spanning receptors and heart function
*H. A. Rockman,
W. J. Koch
& R. J. Lefkowitz*

213 Cardiac channelopathies
E. Marbán

219 New ideas about atrial fibrillation 50 years on
S. Nattel

227 The failing heart
*J. A. Towbin
& N. E. Bowles*

234 Myocardial gene therapy
J. M. Isner

progress

240 Myocyte renewal and ventricular remodelling
*P. Anversa
& B. Nadal-Ginard*

Editor, *Nature*: Philip Campbell
Insights Editor: Lesley Anson
Editorial Assistant: Simon Gibson
Production Editor: Simon Gribbin

Art Director: Majo Xeridat
Diagrams: Ann Thomson
Vicky Askew
Suzanne Coleman

Production Manager: Yvonne Strong
Sponsorship: Emma Jones

Cardiac excitation–contraction coupling

Donald M. Bers

Department of Physiology, Stritch School of Medicine, Loyola University Chicago, 2160 South First Avenue, Maywood, Illinois 60153, USA
(e-mail: dbers@lumc.edu)

Of the ions involved in the intricate workings of the heart, calcium is considered perhaps the most important. It is crucial to the very process that enables the chambers of the heart to contract and relax, a process called excitation–contraction coupling. It is important to understand in quantitative detail exactly how calcium is moved around the various organelles of the myocyte in order to bring about excitation–contraction coupling if we are to understand the basic physiology of heart function. Furthermore, spatial microdomains within the cell are important in localizing the molecular players that orchestrate cardiac function.

Cardiac excitation–contraction coupling is the process from electrical excitation of the myocyte to contraction of the heart (which propels blood out). The ubiquitous second messenger Ca^{2+} is essential in cardiac electrical activity and is the direct activator of the myofilaments, which cause contraction¹. Myocyte mishandling of Ca^{2+} is a central cause of both contractile dysfunction and arrhythmias in pathophysiological conditions².

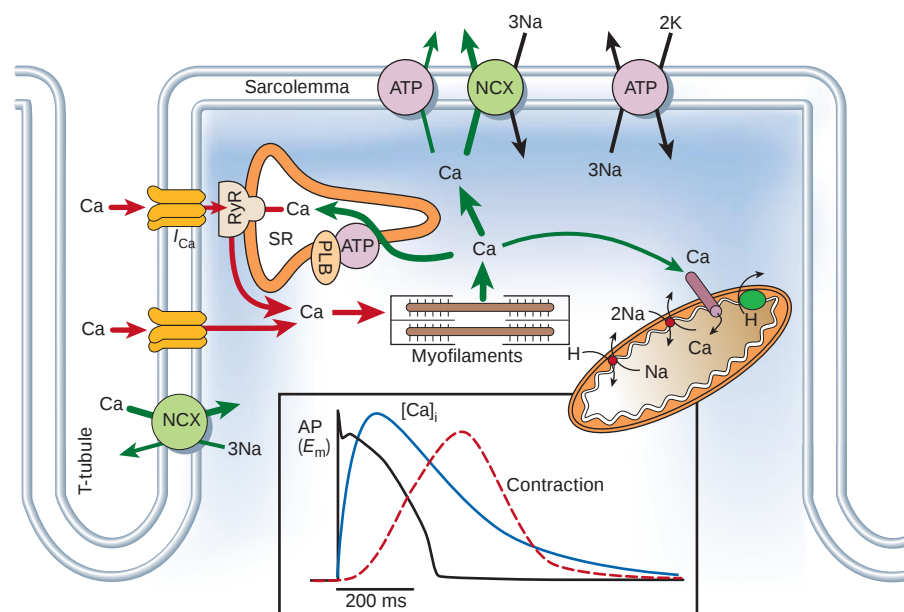
During the cardiac action potential, Ca^{2+} enters the cell through depolarization-activated Ca^{2+} channels as inward Ca^{2+} current (I_{Ca}), which contributes to the action potential plateau (Fig. 1). Ca^{2+} entry triggers Ca^{2+} release from the sarcoplasmic reticulum (SR). The combination of Ca^{2+} influx and release raises the free intracellular Ca^{2+} concentration ($[\text{Ca}^{2+}]_{\text{i}}$), allowing Ca^{2+} to bind to the myofilament protein troponin C, which then switches on the contractile machinery. For relaxation to occur $[\text{Ca}^{2+}]_{\text{i}}$ must decline, allowing Ca^{2+} to dissociate from troponin. This requires

Ca^{2+} transport out of the cytosol by four pathways involving SR Ca^{2+} -ATPase, sarcolemmal $\text{Na}^{+}/\text{Ca}^{2+}$ exchange, sarcolemmal Ca^{2+} -ATPase or mitochondrial Ca^{2+} uniport. Here I discuss the key Ca^{2+} transport systems in cardiac myocytes, how they interact dynamically and how they are regulated. The increasingly important area of local molecular signalling in microdomains will also be addressed.

The role of calcium in contraction and flux balance

Although Ca^{2+} is the switch that activates the myofilaments (the end effectors of excitation–contraction coupling), contraction is graded and depends on $[\text{Ca}^{2+}]_{\text{i}}$ and other factors. Figure 2a shows the amount of total cytosolic $[\text{Ca}^{2+}]$ ($[\text{Ca}^{2+}]_{\text{tot}} = [\text{Ca}^{2+}]_{\text{i}}$ plus bound Ca^{2+}) that must be supplied to and removed from the cytosol during each cardiac beat. Half-maximal activation of contraction requires roughly 70 μmol of Ca^{2+} per litre of cytosol, which would raise $[\text{Ca}^{2+}]_{\text{i}}$ to 600 nM. This ratio of bound:free Ca^{2+} indicates that there is powerful cytosolic Ca^{2+} buffering ($\sim 100:1$)¹.

Figure 1 Ca^{2+} transport in ventricular myocytes. Inset shows the time course of an action potential, Ca^{2+} transient and contraction measured in a rabbit ventricular myocyte at 37 °C. NCX, $\text{Na}^{+}/\text{Ca}^{2+}$ exchange; ATP, ATPase; PLB, phospholamban; SR, sarcoplasmic reticulum.



The development of contraction force depends on $[Ca^{2+}]_i$ and $[Ca^{2+}]_{tot}$ in highly nonlinear relations, as a result of strong myofilament cooperativity with respect to $[Ca^{2+}]_i$ (refs 1,3,4). Moreover, the physiological contraction generates both isometric force (or ventricular pressure) and rapid shortening (to eject blood). There are two main ways to change the strength of cardiac contraction: by altering the amplitude or duration of the Ca^{2+} transient, and by altering the sensitivity of the myofilaments to Ca^{2+} . Myofilament Ca^{2+} sensitivity is enhanced dynamically by stretching the myofilaments (as the heart fills with blood), resulting in a stronger contraction. This is due, in part, to the transverse filament lattice compression that occurs on stretch, which enhances the actin–myosin interaction⁵. This lateral compression is an important autoregulatory mechanism by which the heart adjusts to altered diastolic filling (the classic Frank–Starling response).

Myofilament Ca^{2+} sensitivity is reduced by acidosis, and by elevated phosphate and Mg^{2+} concentrations (all three of which occur during ischaemia). Myofilament Ca^{2+} sensitivity is also reduced by β -adrenergic activation (see below), but enhanced by caffeine and certain inotropic drugs. The rapid kinetics of the Ca^{2+} transient prevent the myofilaments from fully equilibrating with $[Ca^{2+}]_i$ during a normal twitch (especially in the rising phase). Thus, although contractile strength is indicative of underlying Ca^{2+} transients, there is a dynamic interplay between Ca^{2+} and myofilaments during excitation–contraction coupling.

Ca^{2+} must be removed from the cytosol to lower $[Ca^{2+}]_i$ and allow relaxation. This is achieved by several routes, the quantitative importance of which varies between species (ref. 6; and Fig. 2b). In rabbit ventricular myocytes, the SR Ca^{2+} -ATPase pump removes 70% of the activator Ca^{2+} , and Na^+/Ca^{2+} exchange removes 28%, leaving only about 1% each to be removed by the sarcolemmal Ca^{2+} -ATPase and mitochondrial Ca^{2+} uniporter (the last two are collectively referred to as ‘the slow systems’). The amount of Ca^{2+} that leaves the cytosol by entering mitochondria is inconsequential with respect to excitation–contraction coupling, but slow cumulative changes in intra-mitochondrial $[Ca^{2+}]$ can stimulate key dehydrogenases that increase the production of NADH (nicotinamide adenine dinucleotide) and ATP to match increased energetic demands⁷.

The activity of SR Ca^{2+} -ATPase is higher in rat ventricle than in rabbit ventricle (because of a greater concentration of pump molecules)⁸, and Ca^{2+} removal through Na^+/Ca^{2+} exchange is lower, resulting in a balance of 92% for SR Ca^{2+} -ATPase, 7% for Na^+/Ca^{2+} exchange and 1% for the slow systems (Fig. 2b). Analysis in mouse ventricle is quantitatively like rat⁹, whereas the balance of Ca^{2+} fluxes in ferret, dog, cat, guinea-pig and human ventricle are more like rabbit¹. Thus, mouse and rat ventricle (which also show very spike-like action potentials) poorly mimic human with respect to the quantitative balance of cellular Ca^{2+} flux. Moreover, during heart failure in humans and rabbits, functional expression of SR Ca^{2+} -ATPase is reduced and Na^+/Ca^{2+} exchange is increased¹⁰, such that these systems contribute more equally to the decline in $[Ca^{2+}]_i$ (refs 1,2). These changes counterbalance each other with respect to twitch relaxation and $[Ca^{2+}]_i$ decline, leaving it unaltered. But both changes tend to reduce Ca^{2+} content in the SR, limiting SR Ca^{2+} release, and this may be a central cause of systolic contractile deficit in heart failure.

The amount of Ca^{2+} extruded from the cell during twitch relaxation must be the same as the amount of Ca^{2+} entry for each beat, otherwise the cell would gain or lose Ca^{2+} (and would not be in steady state). Indeed, complementary measurements of Ca^{2+} influx and SR Ca^{2+} release during a twitch confirm this expectation in rabbit and rat¹¹. This provides a quantitative framework of dynamic Ca^{2+} fluxes in ventricular myocytes.

Calcium current

Myocytes exhibit two classes of voltage-dependent Ca^{2+} channels (L- and T-type)¹ and the large electrochemical $[Ca^{2+}]$ gradient also

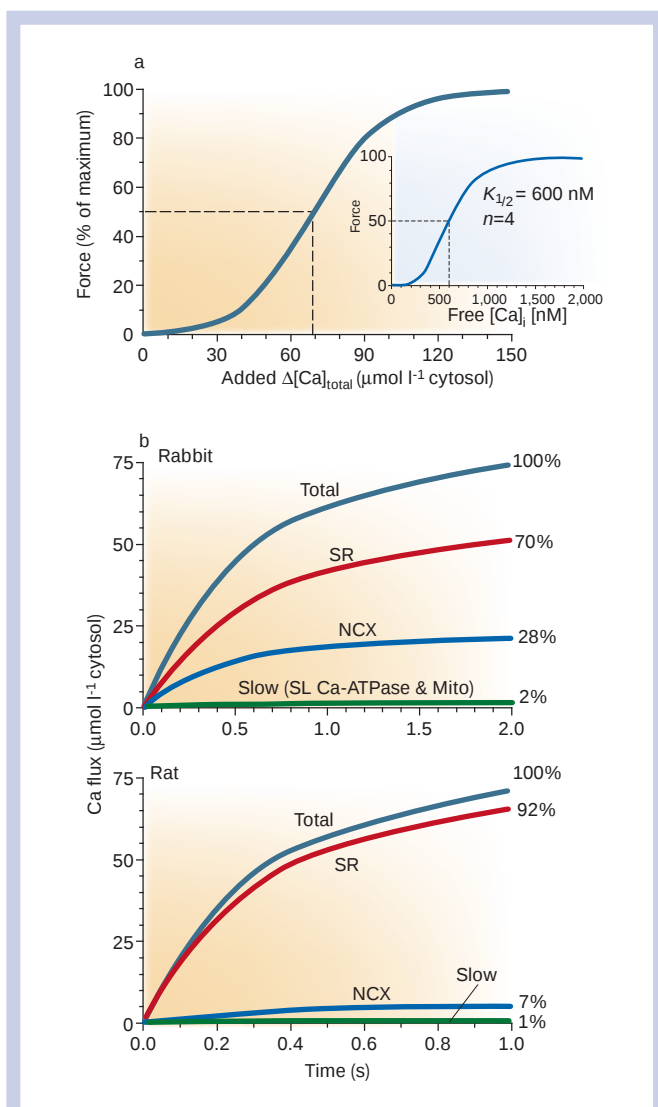
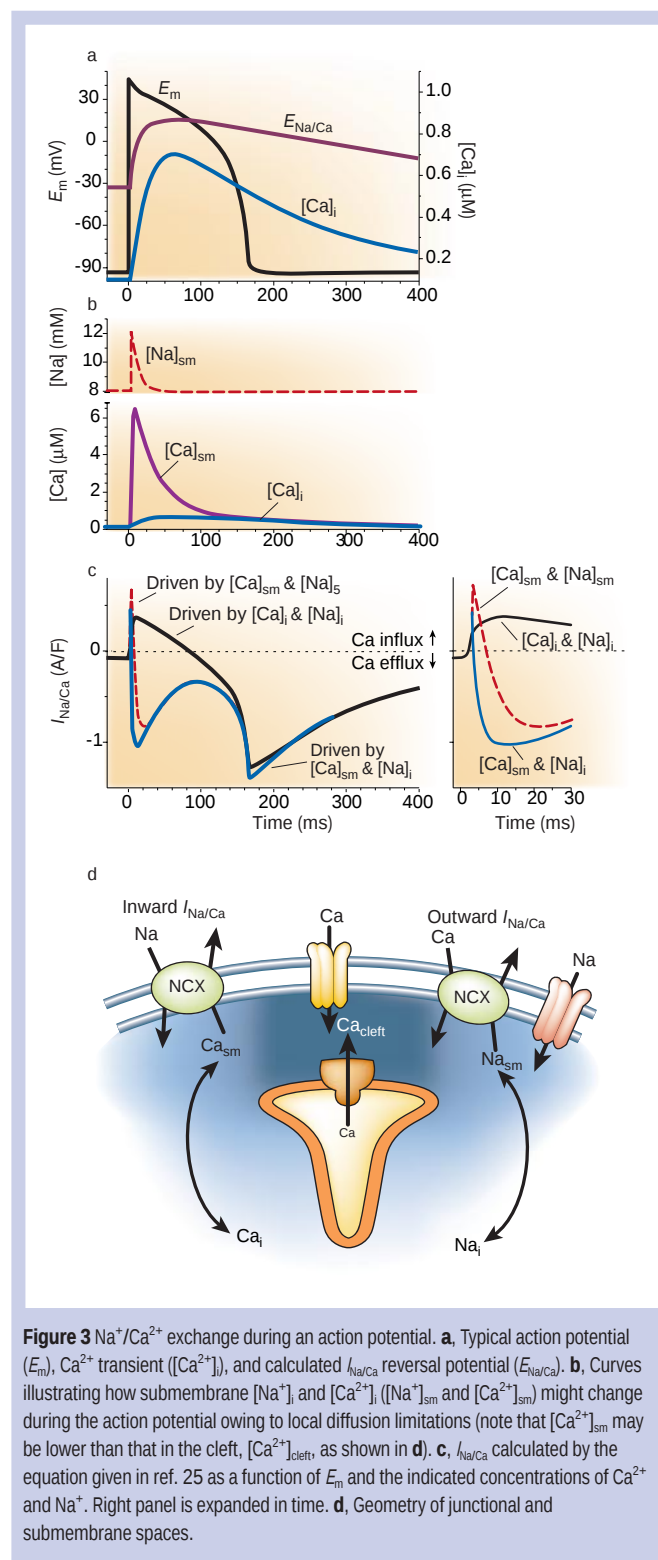


Figure 2 Quantitative Ca^{2+} fluxes during excitation–contraction coupling. **a**, Amount of Ca^{2+} required for contractile activation, assuming a diastolic intracellular Ca^{2+} concentration ($[Ca^{2+}]_i$) of 150 nM and cytosolic Ca^{2+} buffers including troponin C (Ca^{2+} and Ca^{2+}/Mg^{2+} sites), myosin, SR Ca^{2+} -ATPase, calmodulin, ATP, creatine phosphate and sarcolemmal sites¹. Inset shows force as a function of $[Ca^{2+}]_i$ (that is, force is equal to $100/(1 + \{600/[Ca^{2+}]_i\}^4)$; ref. 1). **b**, Integrated Ca^{2+} fluxes during twitch relaxation in rabbit and rat ventricular myocytes. Curves are based on $[Ca^{2+}]_i$ and the $[Ca^{2+}]_i$ dependence of transport rates measured for each system. Percentages are relative contributions to Ca^{2+} removal⁴. SL, sarcolemmal; SR, SR Ca^{2+} -ATPase; Mito, mitochondrial Ca^{2+} uniporter.

drives Ca^{2+} into resting myocytes (at $\sim 1 \mu\text{mol l}^{-1}$ cytosol s^{-1}) by unknown pathways¹². As T-type I_{Ca} is negligible in most ventricular myocytes, I_{Ca} generally refers to the L-type here. I_{Ca} is activated by depolarization, but Ca^{2+} -dependent inactivation at the cytosolic side limits the amount of Ca^{2+} entry during the action potential. This Ca^{2+} -dependent inactivation is a local effect and is mediated by calmodulin bound to the carboxy terminus of the Ca^{2+} channel^{13,14}.

L-type Ca^{2+} channels (dihydropyridine receptors; DHPRs) are located primarily at sarcolemmal–SR junctions where the SR Ca^{2+} release channels (or ryanodine receptors; RyRs) exist¹⁵. During excitation–contraction coupling, SR Ca^{2+} release also contributes to Ca^{2+} -dependent inactivation of I_{Ca} (refs 16,17). Indeed, the total Ca^{2+} influx through I_{Ca} is reduced by about 50% when SR Ca^{2+} release occurs (from 12 to 6 $\mu\text{mol Ca}^{2+} \text{ l}^{-1}$ cytosol)¹⁸. Thus, SR Ca^{2+}



release and I_{Ca} create local negative feedbacks on Ca^{2+} influx. When there is high Ca^{2+} influx or release, further influx of Ca^{2+} is turned off.

The effect of SR Ca^{2+} release on inactivation of I_{Ca} provides a unique bioassay for local cleft $[\text{Ca}^{2+}]_i$ (which may exceed $50 \mu\text{M}$ during excitation–contraction coupling¹⁹). That is, the rapid inactivation of I_{Ca} allows measurement of the SR Ca^{2+} release time course (which peaks in 2–3 ms at 35°C)¹⁸. This confirms the very brief delay expected between I_{Ca} and SR Ca^{2+} release during excitation–contraction coupling²⁰. Thus, the location of Ca^{2+}

channels allows electrophysiological signals that provide spatial information that cannot be assessed by optical fluorescence imaging (owing to the small cleft size).

Calcium influx and efflux

$\text{Na}^+/\text{Ca}^{2+}$ exchange is reversible, with a stoichiometry of three Na^+ ions to one Ca^{2+} ion (but see refs 21,22) that produces an ionic current ($I_{\text{Na/Ca}}$). $\text{Na}^+/\text{Ca}^{2+}$ exchange can extrude Ca^{2+} (as an inward $I_{\text{Na/Ca}}$) or bring Ca^{2+} into the cell (as outward $I_{\text{Na/Ca}}$). $I_{\text{Na/Ca}}$ exhibits a reversal potential that is analogous to those of ion channels ($E_{\text{Na/Ca}} = 3E_{\text{Na}} - 2E_{\text{Ca}}$, where E_{Na} and E_{Ca} are equilibrium potentials for Na^+ and Ca^{2+}). In simpler terms, high $[\text{Ca}^{2+}]_i$ favours Ca^{2+} efflux (inward $I_{\text{Na/Ca}}$), whereas positive membrane potential (E_m) and high $[\text{Na}^+]_i$ favour outward $I_{\text{Na/Ca}}$. Figure 3a shows a typical ventricular action potential, Ca^{2+} transient, and inferred $E_{\text{Na/Ca}}$. At rest $E_m < E_{\text{Na/Ca}}$, so Ca^{2+} extrusion is favoured (inward $I_{\text{Na/Ca}}$; Fig. 3c, black curve). Early in the action potential the E_m exceeds $E_{\text{Na/Ca}}$, which tends to drive Ca^{2+} entry by outward $I_{\text{Na/Ca}}$ (until $E_m = E_{\text{Na/Ca}}$ during repolarization). Note that $E_{\text{Na/Ca}}$ changes because $[\text{Ca}^{2+}]_i$ (and thus E_{Ca}) changes. On repolarization of the action potential, the negative E_m and high $[\text{Ca}^{2+}]_i$ drive a large inward $I_{\text{Na/Ca}}$ and this reflects Ca^{2+} extrusion from the cell.

This simple expectation is complicated by elevations in local submembrane $[\text{Ca}^{2+}]_i$ and $[\text{Na}^+]_i$ ($[\text{Ca}^{2+}]_{\text{sm}}$ and $[\text{Na}^+]_{\text{sm}}$), which are caused by rapid Na^+ and Ca^{2+} fluxes (through I_{Na} , I_{Ca} and SR Ca^{2+} release)^{23,24}. Figure 3b shows possible time courses for $[\text{Ca}^{2+}]_{\text{sm}}$ and $[\text{Na}^+]_{\text{sm}}$ that may be sensed by the $\text{Na}^+/\text{Ca}^{2+}$ exchanger during normal excitation–contraction coupling²⁵. Although $[\text{Ca}^{2+}]_{\text{sm}}$ may not get as high as cleft $[\text{Ca}^{2+}]_i$ during I_{Ca} and SR Ca^{2+} release, the high $[\text{Ca}^{2+}]_{\text{sm}}$ causes $I_{\text{Na/Ca}}$ to become inward very early in the rise of the action potential, such that very little Ca^{2+} enters through $I_{\text{Na/Ca}}$ ($\ll 1 \mu\text{M}$). An I_{Na} -induced rise in $[\text{Na}^+]_{\text{sm}}$ during the action potential might delay the reversal of $I_{\text{Na/Ca}}$ to inward (Fig 3c, red curve), but outward $I_{\text{Na/Ca}}$ would still last for only 4 ms or less (with Ca^{2+} entry still less than $1 \mu\text{M}$). Thus, under physiological conditions $\text{Na}^+/\text{Ca}^{2+}$ exchange works mainly in the Ca^{2+} extrusion mode, driven mostly by the Ca^{2+} transient. The positive E_m during the action potential plateau can, however, limit Ca^{2+} extrusion. This emphasizes again the importance of considering local versus bulk ion concentration (Fig. 3d) as discussed above for inactivation of I_{Ca} .

Although $\text{Na}^+/\text{Ca}^{2+}$ exchange may normally work mainly in the Ca^{2+} efflux mode, the amount of Ca^{2+} influx through $I_{\text{Na/Ca}}$ can be increased greatly if $[\text{Na}^+]_i$ is elevated (for example, by digitalis glycosides that block Na^+/K^+ -ATPase), if SR Ca^{2+} release and/or I_{Ca} is inhibited, or if action potential duration is prolonged²⁶.

Role of the sarcoplasmic reticulum

A high load of Ca^{2+} in the SR directly increases the amount of Ca^{2+} available for release, but also greatly enhances the fraction of SR Ca^{2+} that is released for a given I_{Ca} trigger^{27,28}. This is due, at least in part, to a stimulatory effect of high intra-SR free $[\text{Ca}^{2+}]$ ($[\text{Ca}^{2+}]_{\text{SR}}$) on the open probability of RyRs^{29,30}. This increased RyR sensitivity to $[\text{Ca}^{2+}]_i$ at high $[\text{Ca}^{2+}]_{\text{SR}}$ means that what is often referred to as ‘spontaneous SR Ca^{2+} release’ at high cellular SR Ca^{2+} content ($> 100 \mu\text{mol l}^{-1}$ cytosol) might be considered mechanistically to be triggered by high $[\text{Ca}^{2+}]_{\text{SR}}$ (sometimes in synergy with high $[\text{Ca}^{2+}]_i$). This is the basis of ‘aftercontractions’, transient inward current and delayed after depolarizations that can trigger arrhythmias³.

At moderately low SR Ca^{2+} content, I_{Ca} can fail to induce Ca^{2+} release from the SR^{27,28}. This may help the SR to reload if it becomes relatively depleted. Indeed, low SR Ca^{2+} release allows more Ca^{2+} influx through I_{Ca} (less inactivation) and $\text{Na}^+/\text{Ca}^{2+}$ exchange (less shift towards Ca^{2+} extrusion). A decline in SR Ca^{2+} load, even locally, may contribute dynamically to the turn-off of Ca^{2+} release from the SR during excitation–contraction coupling. SR Ca^{2+} content can be raised by increasing Ca^{2+} influx, decreasing Ca^{2+} efflux, or enhancing Ca^{2+} uptake into the SR (for example, by

adrenergic stimulation or an increase in stimulation frequency, action potential duration, I_{Ca} or $[Na^+]_i$).

Phospholamban is an endogenous inhibitor of the SR Ca^{2+} -ATPase. Phosphorylation of phospholamban by cyclic-AMP-dependent or calmodulin-dependent protein kinases (PKA or CaMKII) relieves this inhibition, allowing faster twitch relaxation and decline of $[Ca^{2+}]_i$. Because the SR Ca^{2+} -ATPase competes better with Na^+/Ca^{2+} exchange, phosphorylation of phospholamban also enhances Ca^{2+} content in the SR. Indeed, targeted gene knockout of phospholamban results in animals with hyperdynamic hearts, with little apparent negative consequence³¹.

The sarcoplasmic reticulum release complex

The RyR is both the SR Ca^{2+} release channel and a scaffolding protein that localizes numerous key regulatory proteins to the junctional complex. These include calmodulin (which can exert Ca^{2+} -dependent modulation of RyR function)³², FK-506 binding protein (FKBP 12.6; which may stabilize RyR gating and also couple the gating of both individual and adjacent RyR tetramers)³³, PKA (which can alter RyR and I_{Ca} gating), phosphatases 1 and 2A (ref. 34) and sorcin (which binds to RyR and DHPR)³⁵. RyRs are also coupled to other proteins at the luminal SR surface (triadin, junctin and calsequestrin)³⁶. These proteins participate in both intra-SR Ca^{2+} buffering and modulation of the Ca^{2+} release process. RyRs are arranged in large organized arrays (up to 200 nm in diameter with more than 100 RyRs) at the junctions between the SR and sarcolemma beneath DHPRs (on the surface and in T-tubules)³⁷. These arrays constitute a large functional Ca^{2+} release complex at the junction (or couplon).

This local functional unit concept is supported by observations of Ca^{2+} sparks or spontaneous local Ca^{2+} transients (Fig. 4). Ca^{2+} sparks reflect the nearly synchronous activation of a cluster of about 6–20 RyRs at a single junction, which is central to the generally accepted local control model of cardiac excitation–contraction coupling^{38–41}. Ca^{2+} sparks are the fundamental units of SR Ca^{2+} release both at rest (rare, stochastic events) and also during excitation–contraction coupling. During excitation–contraction coupling, however, several thousand Ca^{2+} sparks in each cell are synchronized in time by the action potential, such that the local rises in $[Ca^{2+}]_i$ are completely overlapping in time and space (making the Ca^{2+} transient appear spatially uniform). Local Ca^{2+} release events can still be visualized during excitation–contraction coupling by either blocking more than 90% of I_{Ca} (to eliminate overlap of neighbouring events) or by trapping the released Ca^{2+} on exogenous buffers (preventing spatial overlap)^{42–44}.

Resting Ca^{2+} sparks are normally rare and isolated by the space between couplons. But when cellular and SR Ca^{2+} load rise, the exclusively local stochastic cluster behaviour is overcome and Ca^{2+} released at one junction can activate a neighbouring junction (owing to higher $[Ca^{2+}]_{SR}$ and $[Ca^{2+}]_i$) and lead to propagated Ca^{2+} waves and oscillations (Fig. 4d). The waves can propagate the length of a cell (at $\sim 100 \mu m s^{-1}$). When two waves collide they annihilate each other, indicative of RyR refractoriness in the wake of a wave of release.

Activating calcium release

The mechanisms underlying both activation (Fig. 5) and termination of Ca^{2+} release from the SR are controversial. By far the most widely accepted mechanism is Ca^{2+} -induced Ca^{2+} -release (CICR), in particular CICR mediated by the L-type Ca^{2+} channel current ($I_{Ca,L}$). It seems that the opening of one local L-type Ca^{2+} channel in each couplon (and 2–4 Ca^{2+} ions binding to the RyR) is sufficient to activate fully the release process at that couplon¹. In a couplon, neighbouring RyRs are activated either by high local Ca^{2+} ($> 10 \mu M$) or coupled gating between RyRs. This renders an individual couplon all or none, but local $[Ca^{2+}]_i$ declines between couplons, which normally prevents wave propagation. Having more than one Ca^{2+} channel per couplon (10–25 DHPR/100 RyR) creates a safety margin to assure that each couplon will normally fire. Notably, only a

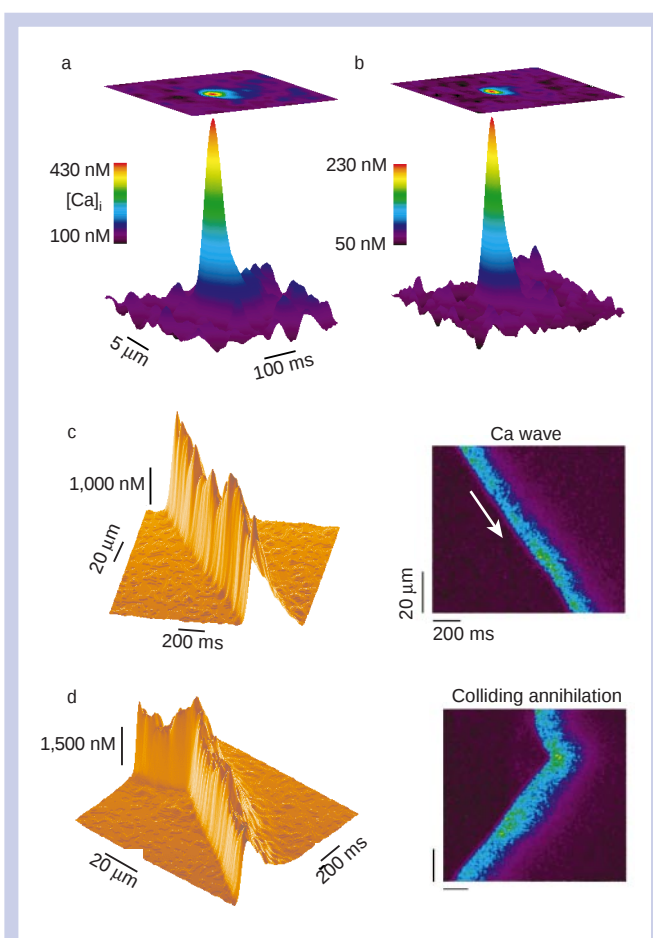


Figure 4 Confocal images of Ca^{2+} sparks and waves. **a,b**, Ca^{2+} sparks rise to a peak in about 10 ms, decline with a time constant of about 20 ms, and have a spatial spread of $2.5 \mu m$ (full-width at half-maximum). Ca^{2+} sparks were recorded in intact and permeabilized mouse ventricular myocytes, respectively. **c,d**, Ca^{2+} waves recorded in intact myocytes at high $[Ca^{2+}]_o$ (extracellular $[Ca^{2+}]$). Top panels in **a** and **b** and right panels in **c** and **d** are longitudinal line scan records (length versus time); the others are surface plots (length versus time versus $[Ca^{2+}]_i$). Local $[Ca^{2+}]_i$ is in pseudocolour in most panels, and the arrow indicates the direction of wave propagation at $93 \mu m s^{-1}$. Figure prepared by Y. Li.

fraction of the L-type Ca^{2+} channels and RyRs in a cell or couplon needs to open to produce the measured Ca^{2+} fluxes¹.

Overwhelming data support this general model as the main excitation–contraction coupling mechanism in heart, but contributions from other pathways have been also been proposed. T-type I_{Ca} ($I_{Ca,T}$) might work like $I_{Ca,L}$, but is nonfunctional in most ventricular myocytes. Even where present (in some Purkinje and atrial cells), a given $I_{Ca,T}$ is very much weaker than $I_{Ca,L}$ in triggering Ca^{2+} release^{45,46}. Thus, T-type Ca^{2+} channels are not preferentially located in junctional regions, and might only be a very minor contributor to excitation–contraction coupling.

Ca^{2+} influx through Na^+/Ca^{2+} exchange has also been proposed to trigger SR Ca^{2+} release in two different ways. First, Na^+ current can raise local $[Na^+]_{sm}$ (see Fig. 3b), causing Ca^{2+} entry through $I_{Na/Ca}$ to trigger SR Ca^{2+} release^{24,47,48}, however, this interpretation has been challenged^{49–51}. Na^+ channels might also be excluded from the junctional cleft¹⁵, making this mechanism less plausible. Second, outward $I_{Na/Ca}$ is activated directly by depolarization (Fig. 3), and can trigger SR Ca^{2+} release and contraction, especially at very positive E_m and when $I_{Ca,L}$ is blocked^{52,53}. But a given Ca^{2+} influx through $I_{Na/Ca}$ is much less effective and slower than $I_{Ca,L}$ in triggering SR Ca^{2+} release⁵⁴, therefore, when both I_{Ca} and $I_{Na/Ca}$ triggers coexist CICR is controlled

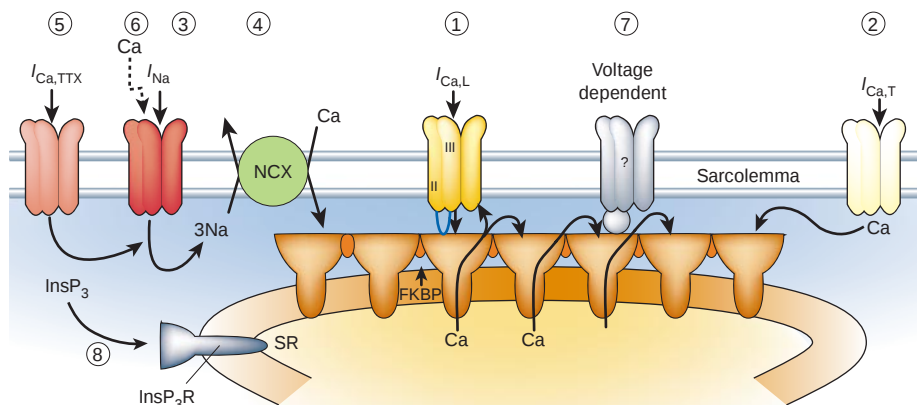


Figure 5 Candidate mechanisms for activation of Ca^{2+} release from the sarcoplasmic reticulum. Numbers refer to the order in which each is discussed in the text. The proposed mechanisms are as follows: 1, Ca^{2+} -induced Ca^{2+} -release (CICR) mediated by L-type Ca^{2+} channel current ($I_{\text{Ca,L}}$); 2, CICR mediated by T-type Ca^{2+} channel current ($I_{\text{Ca,T}}$); 3 and 4, CICR triggered by calcium influx through $\text{Na}^+/\text{Ca}^{2+}$ exchange; 5, CICR

triggered by calcium entry through tetrodotoxin (TTX)-sensitive Ca^{2+} current ($I_{\text{Ca,TTX}}$); 6, Ca^{2+} release mediated by 'slip mode conductance' in which Na^+ channels have altered preference for Ca^{2+} ; 7, voltage-dependent Ca^{2+} release; and 8, Ca^{2+} release triggered by inositol (1,4,5) trisphosphate (InsP_3) through InsP_3 receptors. FKBP, FK-506 binding protein; NCX, $\text{Na}^+/\text{Ca}^{2+}$ exchange.

almost completely by I_{Ca} . $\text{Na}^+/\text{Ca}^{2+}$ exchanger molecules may be largely excluded from the junctional cleft¹⁵. Thus, although outward $I_{\text{Na/Ca}}$ can trigger SR Ca^{2+} release, its physiological role might be to raise $[\text{Ca}^{2+}]_{\text{sm}}$ during the latent time before a particular Ca^{2+} channel opening (or to buoy junctional $[\text{Ca}^{2+}]_i$).

Calcium entry through tetrodotoxin (TTX)-sensitive Ca^{2+} current ($I_{\text{Ca,TTX}}$) has been reported to occur in the absence of $[\text{Na}^+]_o$ (extracellular $[\text{Na}^+]$) and has been attributed to a distinct subpopulation of Na^+ channels^{55,56}. Although $I_{\text{Ca,TTX}}$ might trigger CICR, it seems unlikely that appreciable Ca^{2+} entry occurs at physiological $[\text{Na}^+]_o$. Provocative data suggest that cardiac Na^+ channel selectivity is altered markedly by either β -adrenergic agonists or cardiac glycosides, making Na^+ channels prefer Ca^{2+} over Na^+ (termed 'slip-mode conductance')^{57,58}. This TTX-sensitive Ca^{2+} entry might even trigger SR Ca^{2+} release. The huge change in Na^+ channel selectivity (slip-mode) induced by cAMP or ouabain has eluded detection by other groups and remains controversial^{59–61}. It is unclear whether $I_{\text{Ca,TTX}}$ and slip-mode conductance are related, or if either really contributes to cardiac excitation–contraction coupling³⁹.

Despite overwhelming evidence that Ca^{2+} influx is essential for cardiac excitation–contraction coupling, a few studies have suggested a voltage-dependent Ca^{2+} release that does not require Ca^{2+} influx⁶². Several major concerns have strongly challenged this hypothesis, however, and at this point it is not convincing^{1,39,63}.

Inositol (1,4,5)-trisphosphate (InsP_3) can trigger Ca^{2+} release from smooth muscle SR and endoplasmic reticulum in many cell types, by means of InsP_3 receptors. There are InsP_3 receptors in ventricular myocytes (primarily isoform 2)^{64,65}. Although high concentrations of InsP_3 can cause Ca^{2+} release in cardiac myocytes (particularly atrial cells, which have more InsP_3 receptors), the rate and extent of Ca^{2+} release are very much lower than for CICR, and action potentials are not known to stimulate InsP_3 production⁶⁶. Moreover, cardiac α_1 -adrenergic and muscarinic agonists increase production of InsP_3 and contractile force^{67,68}, but this inotropic effect is mediated mainly by protein kinase C rather than InsP_3 (refs 69,70).

These neurohumoral agents also stimulate hypertrophic gene transcription in a Ca^{2+} -dependent manner⁷¹. The InsP_3 produced may bind to perinuclear InsP_3 receptors (G. A. Mignery and D.M.B., unpublished observation), raising local $[\text{Ca}^{2+}]_i$ in this

discrete microdomain, thereby activating calmodulin-dependent transcriptional regulation pathways (for example, CaMKII or calcineurin). This speculation is untested, but could represent a local control of excitation–transcription coupling, which is distinct from that involved in excitation–contraction coupling. On balance, InsP_3 has, at most, a minor modulatory role in cardiac excitation–contraction coupling, but may serve other spatially and functionally discrete roles.

In summary, cardiac SR Ca^{2+} release occurs mainly through CICR, and $I_{\text{Ca,L}}$ is the dominant Ca^{2+} source. The other mechanisms shown in Fig. 5 may slightly modulate Ca^{2+} release or be redundant back-up systems. A central limitation for some mechanisms is that the Ca^{2+} flux is not as spatially focused on the RyR as it is for $I_{\text{Ca,L}}$.

Terminating calcium release

Ca^{2+} -induced Ca^{2+} -release is inherently a positive-feedback mechanism, but its turn-off is essential for diastolic refilling of the heart. So what turns release off? Three possibilities include local depletion of SR Ca^{2+} , RyR inactivation (or adaptation), and stochastic attrition^{44,72,73}. Stochastic attrition means that if the L-type Ca^{2+} channels and all RyRs in a junction happen to be closed simultaneously (as channels gate stochastically), then local $[\text{Ca}^{2+}]_i$ will fall very rapidly and interrupt the otherwise regenerative release. This might work for 1 DHPR and 1–2 RyRs, but with more realistic numbers of channels it becomes too unlikely that they will all close at once. The coupled gating of RyRs⁷⁴ might overcome this limitation, especially if most RyRs in a spark site are really coupled that way (but this remains to be tested).

Local depletion of SR Ca^{2+} cannot explain completely the turn-off of release, because very long lasting Ca^{2+} sparks are observed that do not decline with time (>200 ms)^{38,75}. Thus, diffusion from other regions of the SR can limit local Ca^{2+} depletion in the SR. During a global Ca^{2+} transient, however, the whole $[\text{Ca}^{2+}]_{\text{SR}}$ declines. Because $[\text{Ca}^{2+}]_{\text{SR}}$ modulates RyR gating, $[\text{Ca}^{2+}]_{\text{SR}}$ depletion might contribute to shutting-off global SR Ca^{2+} release during a twitch. But, as stated above, this cannot explain fully why Ca^{2+} sparks turn off or why SR Ca^{2+} release terminates.

Two types of RyR inactivation have been reported (and both depend on $[\text{Ca}^{2+}]_i$). One is an absorbing inactivation (as in Na^+ channels), in which the RyR is unavailable for reopening until it recovers^{44,76–78}. The second type is called adaptation, in which the RyR

after activating relaxes to a lower open probability, but can still be reactivated by a higher $[Ca^{2+}]_i$ (refs 79,80). Whether only one of these is functionally relevant remains controversial, and few cellular studies have addressed this unequivocally. But there is clearly some refractoriness in cellular and local events of SR Ca^{2+} release^{44,75}. Recovery of RyR availability occurs with two time constants: one fast (100–300 ms) and one very slow (several seconds). Inactivation of RyRs may be important in minimizing inappropriate SR Ca^{2+} release events between heartbeats. In summary, it seems that both RyR inactivation and partial luminal depletion of SR Ca^{2+} (to reduce RyR opening) both contribute to the turn-off of release. Coupled gating of RyRs (so many gate as one) may also mean that a variant of stochastic attrition also contributes.

Modulation of calcium by sympathetic activation

Physiological sympathetic stimulation of the heart through β -adrenergic receptors increases developed contractions (inotropy) and accelerates relaxation (lusitropy) and $[Ca^{2+}]_i$ decline (Fig. 6). β -Adrenergic receptor stimulation activates a GTP-binding protein (G_s), which stimulates adenylyl cyclase to produce cAMP, which in turn activates PKA. This kinase phosphorylates several proteins related to excitation–contraction coupling (phospholamban, L-type Ca^{2+} channels, RyR, troponin I and myosin binding protein C).

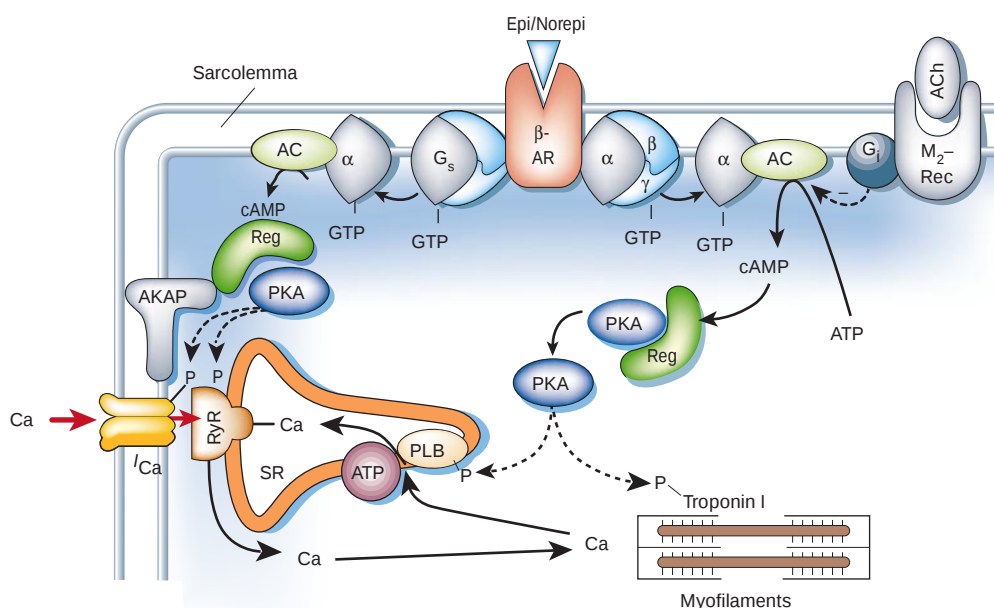
The lusitropic effect of PKA is mediated by phosphorylation of phospholamban and troponin I, which speed up SR Ca^{2+} re-uptake and dissociation of Ca^{2+} from the myofilaments, respectively. But phosphorylation of phospholamban is by far the dominant mechanism for both the lusitropic effect and accelerating the decline in $[Ca^{2+}]_i$ (ref. 81). The faster SR Ca^{2+} uptake also contributes to increasing the SR Ca^{2+} content. The inotropic effect of PKA activation is mediated by the combination of increased I_{Ca} and greater availability of SR Ca^{2+} . This synergistic combination greatly enhances Ca^{2+} transient amplitude, and more than offsets the reduction in myofilament Ca^{2+} sensitivity (caused by troponin I phosphorylation, which by itself would reduce force). The depressant of PKA on the myofilaments seem to be completely attributable to phosphorylation of troponin I (versus myosin-binding protein C), because substitution of troponin I with a non-phosphorylatable troponin I abolishes the myofilament effects of PKA⁸².

PKA can also modulate the open probability of RyR channels. In isolated single-channel recordings, PKA increased initial RyR opening during an abrupt $[Ca^{2+}]_i$ rise, but decreased the steady-state open probability at a given $[Ca^{2+}]_i$ (ref. 80). In contrast, Marx *et al.*³³ found that PKA enhanced the steady-state open probability of single RyRs in bilayers, and attributed this to the displacement of FKBP-12.6 from the RyR. Moreover, they found that RyRs were hyperphosphorylated in heart failure, which could cause a diastolic leak of SR Ca^{2+} and contribute to the reduced SR Ca^{2+} content in heart failure (see above). But in more intact cellular systems, no effect of PKA-dependent RyR phosphorylation could be detected on resting SR Ca^{2+} leak (as Ca^{2+} sparks) in the absence of phospholamban (with unchanged SR Ca^{2+} load)⁸³. Phosphorylation of RyRs may also alter the intrinsic responsiveness of SR Ca^{2+} release to an I_{Ca} trigger signal, but results concerning this have been mixed, showing an increase, decrease and lack of change^{84–86}. Thus, whether PKA-dependent phosphorylation alters RyR behaviour during rest or excitation–contraction coupling remains controversial. This process is particularly challenging to measure in intact cells, because increases in I_{Ca} and in SR Ca^{2+} uptake make isolation of intrinsic RyR effects difficult.

Eisner *et al.*⁸⁷ have also argued that, because of autoregulation, altered systolic gating properties of RyRs in intact cells alone exert only transitory effects on Ca^{2+} transient amplitude. That is, abrupt increases in RyR opening or fractional SR Ca^{2+} release cause greater Ca^{2+} extrusion through Na^+/Ca^{2+} exchange at the first beat, thereby decreasing SR Ca^{2+} available for the next beat. In the steady state, this lower SR Ca^{2+} content offsets the increased fractional SR Ca^{2+} release such that Ca^{2+} transients are almost unchanged. However, enhanced diastolic leak of SR Ca^{2+} might still contribute to reduced SR Ca^{2+} load and systolic function in heart failure.

Local signalling is also important in the β -adrenergic receptor cascade. L-type Ca^{2+} channels co-assemble with β_2 -adrenergic receptors, G_s , adenylyl cyclase, PKA and phosphatase 2A (at least in brain)⁸⁸. The cardiac RyR serves as both a PKA target and a scaffolding protein (where PKA and phosphatases 1 and 2A are all bound to the RyR through anchoring proteins)³⁴. The close physical proximity may be functionally essential⁸⁹. The activation of β_1 -adrenergic receptors in ventricular myocytes produces robust inotropic and lusitropic effects, paralleled by phosphorylation of Ca^{2+} channels,

Figure 6 β -Adrenergic receptor activation and phosphorylation targets relevant to excitation–contraction coupling. AC, adenylyl cyclase; ACh, acetylcholine; AKAP, A kinase anchoring protein; β -AR, β -adrenergic receptor; M_2 -Rec, M_2 -muscarinic receptor; PLB, phospholamban; Reg, PKA regulatory subunit; SR, sarcoplasmic reticulum.



phospholamban and troponin I. By contrast, the activation of β_2 -adrenergic receptors may be more restricted to I_{Ca} enhancement⁹⁰, and β_2 -adrenergic receptors are located almost exclusively in specialized sarcolemmal invaginations called caveolae (versus β_1 -adrenergic receptors, which are largely non-caveolar)⁹¹.

Activation of other G-protein-coupled receptors that greatly stimulate cAMP production (for example, prostaglandin E, histamine, glucagon-like peptide-1) produce little or no inotropic effect (as compared with β_1 -adrenergic receptors)⁹². Thus, the functionally important levels of cAMP, activated PKA, phosphatase and phosphodiesterase (which breaks down cAMP) are those very near to that of the target protein. The total cellular concentration of cAMP might be irrelevant to key regulatory pathways, except as an overflow from local cAMP-mediated signal transduction. However, if this is true, it is less clear how targeting would practically work for phospholamban and troponin I phosphorylation (as compared with I_{Ca} or RyR). This would require very high amounts of the various anchoring and signalling proteins, because troponin I and phospholamban are present at 50 μ M or higher concentrations and are dispersed widely in the cell.

Relative receptor locations can also regulate this signalling cascade. For example, M_2 -muscarinic receptor activation can either decrease or increase concentrations of cAMP levels, depending on whether they were produced by β_1 - or β_2 -adrenergic receptors, respectively⁹³. This may be due in part to the relative exclusion of M_2 -muscarinic receptors from caveolae. Thus, the location of receptors and their signalling cascade components can selectively determine function.

Implications for calcium handling

Calcium in cardiac myocyte is in a dynamic yet delicate balance, created by multiple interacting cellular systems that can be tuned by physiological modulators. It is also clear that we must think increasingly in terms of microdomains and local control, without losing perspective on the integrative framework in which these domains function. RyR and I_{Ca} are both responsible for, but also controlled by, the local cleft $[Ca^{2+}]_i$ (which may differ greatly from the $[Ca^{2+}]_{sm}$ that controls Na^+/Ca^{2+} exchange). Key regulatory pathways (for example, β -adrenergic receptors, calmodulin and possibly Ca^{2+} -dependent transcription) also exhibit local functional coupling in microdomains. These various signalling domains surely overlap spatially and functionally.

It will be important to develop new experimental tools to assess how the key signalling molecules (including Ca^{2+}) interact functionally and are targeted to the appropriate microdomains. Future studies will need to clarify how cells distinguish between Ca^{2+} involved in excitation-contraction coupling and transcriptional regulation. We may also have to start thinking more stochastically about local reactions. For example, at resting $[Ca^{2+}]_i$ there is less than one free Ca^{2+} ion in an entire junctional cleft, making the concept of collision probability more meaningful than concentration. Thus, much challenging work lies ahead if we are to understand the physiological functions of many of these processes *in situ*, particularly with respect to signalling in microdomains, as well as the pathophysiological and therapeutic implications. □

- Bers, D. M. *Excitation-Contraction Coupling and Cardiac Contractile Force* edn 2 (Kluwer Academic, Dordrecht, Netherlands, 2001).
- Pogwizd, S. M., Schlotthauer, K., Li, L., Yuan, W. & Bers, D. M. Arrhythmogenesis and contractile dysfunction in heart failure: roles of sodium-calcium exchange, inward rectifier potassium current and residual β -adrenergic responsiveness. *Circ. Res.* **88**, 1159–1167 (2001).
- Solaro, R. J. & Rarick, H. M. Troponin and tropomyosin—proteins that switch on and tune in the activity of cardiac myofibrils. *Circ. Res.* **83**, 471–480 (1998).
- Moss, R. L. & Buck, S. H. in *Handbook of Physiology* (eds Page, E., Fozzard, H. A. & Solaro, R. J.) 420–454 (Oxford Univ. Press, New York, 2001).
- Fukuda, N., Sasaki, D., Ishiwata, S. & Kurihara, S. Length dependence of tension generation in rat skinned cardiac muscle. *Circ. Res.* **104**, 1639–1645 (2001).
- Bassani, J. W. M., Bassani, R. A. & Bers, D. M. Relaxation in rabbit and rat cardiac cells: species-dependent differences in cellular mechanisms. *J. Physiol.* **476**, 279–293 (1994).
- Brandes, R. & Bers, D. M. Intracellular Ca^{2+} increases the mitochondrial NADH concentration during elevated work in intact cardiac muscle. *Circ. Res.* **80**, 82–87 (1997).

- Hove-Madsen, L. & Bers, D. M. Sarcoplasmic reticulum Ca^{2+} uptake and thapsigargin sensitivity in permeabilized rabbit and rat ventricular myocytes. *Circ. Res.* **73**, 820–828 (1993).
- Li, L., Chu, G., Kranias, E. G. & Bers, D. M. Cardiac myocyte calcium transport in phospholamban knockout mouse: relaxation and endogenous CaMKII effects. *Am. J. Physiol.* **274**, H1335–H1347 (1998).
- Hasenfuss, G. Alterations of calcium-regulatory proteins in heart failure. *Cardiovasc. Res.* **37**, 279–289 (1998).
- Delbridge, L. M., Bassani, J. W. M. & Bers, D. M. Steady-state twitch Ca^{2+} fluxes and cytosolic Ca^{2+} buffering in rabbit ventricular myocytes. *Am. J. Physiol.* **270**, C192–C199 (1996).
- Trafford, A. W., Díaz, M. E., Negretti, N. & Eisner, D. A. Enhanced Ca^{2+} current and decreased Ca^{2+} efflux restore sarcoplasmic reticulum Ca^{2+} content after depletion. *Circ. Res.* **81**, 477–484 (1997).
- Peterson, B. Z., DeMaria, C. D. & Yue, D. T. Calmodulin is the Ca^{2+} sensor for Ca^{2+} -dependent inactivation of L-type calcium channels. *Neuron* **22**, 549–558 (1999).
- Zühlke, R. D., Pitt, G. S., Deisseroth, K., Tsien, R. W. & Reuter, H. Calmodulin supports both inactivation and facilitation of L-type calcium channels. *Nature* **399**, 159–162 (1999).
- Scriven, D. R. L., Dan, P. & Moore, E. D. W. Distribution of proteins implicated in excitation-contraction coupling in rat ventricular myocytes. *Biophys. J.* **79**, 2682–2691 (2000).
- Sipido, K. R., Callewaert, G. & Carmeliet, E. Inhibition and rapid recovery of Ca^{2+} current during Ca^{2+} release from sarcoplasmic reticulum in guinea pig ventricular myocytes. *Circ. Res.* **76**, 102–109 (1995).
- Sham, J. S. K. et al. Termination of Ca^{2+} release by a local inactivation of ryanodine receptors in cardiac myocytes. *Proc. Natl Acad. Sci. USA* **95**, 15096–15101 (1998).
- Puglisi, J. L., Yuan, W., Bassani, J. W. M. & Bers, D. M. Ca^{2+} influx through Ca^{2+} channels in rabbit ventricular myocytes during action potential clamp: influence of temperature. *Circ. Res.* **85**, e7–e16 (1999).
- Langer, G. A. & Peskoff, A. Calcium concentration and movement in the diadic cleft space of the cardiac ventricular cell. *Biophys. J.* **70**, 1169–1182 (1996).
- Zahradnicková, A., Zahradník, I., Györke, I. & Györke, S. Rapid activation of the cardiac ryanodine receptor by submillisecond calcium stimuli. *J. Gen. Physiol.* **114**, 787–798 (1999).
- Fujioka, Y., Komeda, M. & Matsuoka, S. Stoichiometry of Na^+ - Ca^{2+} exchange in inside-out patches excised from guinea-pig ventricular myocytes. *J. Physiol.* **523**, 339–351 (2000).
- Egger, M. & Niggli, E. Paradoxical block of the Na^+ - Ca^{2+} exchanger by extracellular protons in guinea-pig ventricular myocytes. *J. Physiol.* **523**, 353–366 (2000).
- Trafford, A. W., Díaz, M. E., O'Neill, S. C. & Eisner, D. A. Comparison of subsarcolemmal and bulk calcium concentration during spontaneous calcium release in rat ventricular myocytes. *J. Physiol.* **488**, 577–586 (1995).
- Leblanc, N. & Hume, J. R. Sodium current-induced release of calcium from cardiac sarcoplasmic reticulum. *Science* **248**, 372–376 (1990).
- Weber, C. R., Piacentino, V. III, Ginsburg, K. S., Houser, S. R. & Bers, D. M. Na/Ca exchange current and submembrane $[Ca]$ during cardiac action potential. *Circ. Res.* (in press).
- Dipla, K., Mattiello, J. A., Margulies, K. B., Jeevanandam, V. & Houser, S. R. The sarcoplasmic reticulum and the Na^+/Ca^{2+} exchanger both contribute to the Ca^{2+} transient of failing human ventricular myocytes. *Circ. Res.* **84**, 435–444 (1999).
- Bassani, J. W. M., Yuan, W. & Bers, D. M. Fractional SR Ca release is regulated by trigger Ca and SR Ca content in cardiac myocytes. *Am. J. Physiol.* **268**, C1313–C1319 (1995).
- Shannon, T. R., Ginsburg, K. S. & Bers, D. M. Potentiation of fractional SR Ca release by total and free intra-SR Ca concentration. *Biophys. J.* **78**, 334–343 (2000).
- Sitsapesan, R. & Williams, A. J. Regulation of the gating of the sheep cardiac sarcoplasmic reticulum Ca^{2+} -release channel by luminal Ca^{2+} . *J. Membr. Biol.* **137**, 215–226 (1994).
- Lukyanenko, V., Györke, I. & Györke, S. Regulation of calcium release by calcium inside the sarcoplasmic reticulum in ventricular myocytes. *Pflügers Arch.* **432**, 1047–1054 (1996).
- Brittsan, A. G. & Kranias, E. G. Phospholamban and cardiac contractile function. *J. Mol. Cell. Cardiol.* **32**, 2131–2139 (2000).
- Fruen, B. R., Bardy, J. M., Byrem, T. M., Strasburg, G. M. & Louis, C. F. Differential Ca^{2+} sensitivity of skeletal and cardiac muscle ryanodine receptors in the presence of calmodulin. *Am. J. Physiol.* **279**, C724–C733 (2000).
- Marx, S. O. et al. PKA phosphorylation dissociates FKBP12.6 from the calcium release channel (ryanodine receptor): defective regulation in failing hearts. *Cell* **101**, 365–376 (2000).
- Marx, S. O. et al. Phosphorylation-dependent regulation of ryanodine receptors: a novel role for leucine/isoleucine zippers. *J. Cell Biol.* **153**, 699–708 (2001).
- Meyers, M. B. et al. Sorcin associates with the pore-forming subunit of voltage-dependent L-type Ca^{2+} channels. *J. Biol. Chem.* **273**, 18930–18935 (1998).
- Zhang, L., Kelley, J., Schmeisser, G., Kobayashi, Y. M. & Jones, L. R. Complex formation between junctin, triadin, calsequestrin, and the ryanodine receptor: Proteins of the cardiac junctional sarcoplasmic reticulum membrane. *J. Biol. Chem.* **272**, 23389–23397 (1997).
- Franzini-Armstrong, C., Protasi, F. & Ramesh, V. Shape, size, and distribution of Ca^{2+} release units and couplons in skeletal and cardiac muscles. *Biophys. J.* **77**, 1528–1539 (1999).
- Cheng, H., Lederer, W. J. & Cannell, M. B. Calcium sparks: elementary events underlying excitation-contraction coupling in heart muscle. *Science* **262**, 740–744 (1993).
- Wier, W. G. & Balke, C. W. Ca^{2+} release mechanisms, Ca^{2+} sparks, and local control of excitation-contraction coupling in normal heart muscle. *Circ. Res.* **85**, 770–776 (1999).
- Bridge, J. H. B., Eshler, P. R. & Cannell, M. B. Properties of Ca^{2+} sparks evoked by action potentials in mouse ventricular myocytes. *J. Physiol.* **518**, 469–478 (1999).
- Lukyanenko, V. et al. Inhibition of Ca^{2+} sparks by ruthenium red in permeabilized rat ventricular myocytes. *Biophys. J.* **79**, 1273–1284 (2000).
- Cannell, M. B., Cheng, H. & Lederer, W. J. The control of calcium release in heart muscle. *Science* **268**, 1045–1049 (1995).
- López-López, J. R., Shacklock, P. S., Balke, C. W. & Wier, W. G. Local calcium transients triggered by single L-type calcium channel currents in cardiac cells. *Science* **268**, 1042–1045 (1995).
- Sham, J. S. K. et al. Termination of Ca^{2+} release by a local inactivation of ryanodine receptors in cardiac myocytes. *Proc. Natl Acad. Sci. USA* **95**, 15096–15101 (1998).
- Sipido, K. R., Carmeliet, E. & van de Werf, F. T-type Ca^{2+} current as a trigger for Ca^{2+} release from the sarcoplasmic reticulum in guinea-pig ventricular myocytes. *J. Physiol.* **508**, 439–451 (1998).
- Zhou, Z. F. & January, C. T. Both T- and L-type Ca^{2+} channels can contribute to excitation-contraction coupling in cardiac Purkinje cells. *Biophys. J.* **74**, 1830–1839 (1998).
- Levesque, P. C., Leblanc, N. & Hume, J. R. Release of calcium from guinea pig cardiac sarcoplasmic reticulum induced by sodium-calcium exchange. *Cardiovasc. Res.* **28**, 370–378 (1994).

48. Lipp, P. & Niggli, E. Sodium current-induced calcium signals in isolated guinea-pig ventricular myocytes. *J. Physiol.* **474**, 439–446 (1994).
49. Sham, J. S. K., Cleemann, L. & Morad, M. Gating of the cardiac Ca^{2+} release channel: the role of Na^{+} current and Na^{+} - Ca^{2+} exchange. *Science* **255**, 850–853 (1992).
50. Bouchard, R. A., Clark, R. B. & Giles, W. R. Role of sodium-calcium exchange in activation of contraction in rat ventricle. *J. Physiol.* **472**, 391–413 (1993).
51. Sipido, K. R., Carmeliet, E. & Pappano, A. Na^{+} current and Ca^{2+} release from the sarcoplasmic reticulum during action potentials in guinea-pig ventricular myocytes. *J. Physiol.* **489**, 1–17 (1995).
52. Levi, A. J., Spitzer, K. W., Kohmoto, O. & Bridge, J. H. B. Depolarization-induced Ca entry via Na - Ca exchange triggers SR release in guinea pig cardiac myocytes. *Am. J. Physiol.* **266**, H1422–H1433 (1994).
53. Litwin, S. E., Li, J. & Bridge, J. H. B. Na - Ca exchange and the trigger for sarcoplasmic reticulum Ca release: studies in adult rabbit ventricular myocytes. *Biophys. J.* **75**, 359–371 (1998).
54. Sipido, K. R., Maes, M. & van de Werf, F. Low efficiency of Ca^{2+} entry through the Na^{+} - Ca^{2+} exchanger as trigger for Ca^{2+} release from the sarcoplasmic reticulum—a comparison between L-type Ca^{2+} current and reverse-mode Na^{+} - Ca^{2+} exchange. *Circ. Res.* **81**, 1034–1044 (1997).
55. Lemaire, S., Piot, C., Seguin, J., Nargeot, J. & Richard, S. Tetrodotoxin-sensitive Ca^{2+} and Ba^{2+} currents in human atrial cells. *Recept. Channels* **3**, 71–81 (1995).
56. Aggarwal, R., Shorofsky, S. R., Goldman, L. & Balke, C. W. Tetrodotoxin-blockable calcium currents in rat ventricular myocytes; a third type of cardiac cell sodium current. *J. Physiol.* **505**, 353–369 (1997).
57. Santana, L. F., Gómez, A. M. & Lederer, W. J. Ca^{2+} flux through promiscuous cardiac Na^{+} channels: slip-mode conductance. *Science* **279**, 1027–1033 (1998).
58. Cruz, J. D. S. *et al.* Whether “slip-mode conductance” occurs. *Science* **284**, 711a (1999).
59. Nuss, H. B. & Marbán, E. Whether “slip-mode conductance” occurs. *Science* **284**, 711a (1999).
60. Chandra, R., Chauhan, V. S., Starmer, C. F. & Grant, A. O. β -Adrenergic action on wild-type and KPQ mutant human cardiac Na^{+} channels: shift in gating but no change in Ca^{2+} : Na^{+} selectivity. *Cardiovasc. Res.* **42**, 490–502 (1999).
61. DelPrincipe, F., Egger, M., Niggli, E. L-type Ca^{2+} current as the predominant pathway of Ca^{2+} entry during I_{Na} activation in β -stimulated cardiac myocytes. *J. Physiol.* **527**, 455–466 (2000).
62. Ferrier, G. R. & Howlett, S. E. Cardiac excitation–contraction coupling: role of membrane potential in regulation of contraction. *Am. J. Physiol. (Heart Circ. Physiol.)* **280**, H1928–H1944 (2001).
63. Piacentino, V. III, Dipla, K., Gaughan, J. P. & Houser, S. R. Voltage-dependent Ca^{2+} release from the SR of feline ventricular myocytes is explained by Ca^{2+} -induced Ca^{2+} release. *J. Physiol.* **523**, 533–548 (2000).
64. Perez, P. J., Ramos-Franco, J., Fill, M. & Mignery, G. A. Identification and functional reconstitution of the type 2 inositol 1,4,5-trisphosphate receptor from ventricular cardiac myocytes. *J. Biol. Chem.* **272**, 23961–23969 (1997).
65. Lipp, P. *et al.* Functional InsP_3 receptors that may modulate excitation-contraction coupling in the heart. *Curr. Biol.* **10**, 939–942 (2000).
66. Kentish, J. C. *et al.* Calcium release from cardiac sarcoplasmic reticulum induced by photorelease of calcium or $\text{Ins}(1,4,5)\text{P}_3$. *Am. J. Physiol.* **258**, H610–H615 (1990).
67. Brown, J. H. & Jones, L. G. in *Phosphoinositides and Receptor Mechanisms* (ed. Putney, J. W. Jr) 245–270 (Alan R. Liss, New York, 1986).
68. Poggioli, J., Sulpice, J. C. & Vassort, G. Inositol phosphate production following α_1 -adrenergic, muscarinic, or electrical stimulation in isolated rat heart. *FEBS Lett.* **206**, 292–298 (1986).
69. Endoh, M. Cardiac α_1 -adrenoceptors that regulate contractile function: subtypes and subcellular signal transduction mechanisms. *Neurochem. Res.* **21**, 217–229 (1996).
70. Gambassi, G., Spurgeon, H. A., Ziman, B. D., Lakatta, E. G. & Capogrossi, M. C. Opposing effects of α_1 -adrenergic receptor subtypes on Ca^{2+} and pH homeostasis in rat cardiac myocytes. *Am. J. Physiol.* **274**, H1152–H1162 (1998).
71. Ramirez, M. T., Zhao, X. L., Schulman, H. & Brown, J. H. The nuclear δ_9 isoform of Ca^{2+} /calmodulin-dependent protein kinase II regulates atrial natriuretic factor gene expression in ventricular myocytes. *J. Biol. Chem.* **272**, 31203–31208 (1997).
72. Stern, M. D. Theory of excitation–contraction coupling in cardiac muscle. *Biophys. J.* **63**, 497–517 (1992).
73. Lukyanenko, V., Wiesner, T. F. & Györke, S. Termination of Ca^{2+} release during Ca^{2+} sparks in rat ventricular myocytes. *J. Physiol.* **507**, 667–677 (1998).
74. Marx, S. O. *et al.* Coupled gating between cardiac calcium release channels (ryanodine receptors). *Circ. Res.* **88**, 1151–1158 (2001).
75. Satoh, H., Blatter, L. A. & Bers, D. M. Effects of $[\text{Ca}^{2+}]_i$, SR Ca^{2+} load, and rest on Ca^{2+} spark frequency in ventricular myocytes. *Am. J. Physiol.* **272**, H657–H668 (1997).
76. Fabiato, A. Rapid ionic modifications during the aequorin-detected calcium transient in a skinned canine cardiac Purkinje cell. *J. Gen. Physiol.* **85**, 189–246 (1985).
77. Schiefer, A., Meissner, G. & Isenberg, G. Ca^{2+} activation and Ca^{2+} inactivation of canine reconstituted cardiac sarcoplasmic reticulum Ca^{2+} -release channels. *J. Physiol.* **489**, 337–348 (1995).
78. Sitsapesan, R., Montgomery, R. A. P. & Williams, A. J. New insights into the gating mechanisms of cardiac ryanodine receptors revealed by rapid changes in ligand concentration. *Circ. Res.* **77**, 765–772 (1995).
79. Györke, S. & Fill, M. Ryanodine receptor adaptation: control mechanism of Ca^{2+} -induced Ca^{2+} release in heart. *Science* **260**, 807–809 (1993).
80. Valdivia, H. H., Kaplan, J. H., Ellis-Davies, G. C. R. & Lederer, W. J. Rapid adaptation of cardiac ryanodine receptors: modulation by Mg^{2+} and phosphorylation. *Science* **267**, 1997–2000 (1995).
81. Li, L., DeSantiago, J., Chu, G., Kranias, E. G. & Bers, D. M. Phosphorylation of phospholamban and troponin I in β -adrenergic-induced acceleration of cardiac relaxation. *Am. J. Physiol.* **278**, H769–H779 (2000).
82. Kentish, J. C. *et al.* Phosphorylation of troponin I by protein kinase A accelerates relaxation and crossbridge cycle kinetics in mouse ventricular muscle. *Circ. Res.* **88**, 1059–1065 (2001).
83. Li, Y. & Bers, D. M. Protein kinase A phosphorylation of the ryanodine receptor does not alter Ca sparks in permeabilized mouse ventricular myocyte. *Circulation* **104**, II-131 (2001).
84. Viatchenko-Karpinski, S. & Györke, S. Modulation of the Ca^{2+} -induced Ca^{2+} release cascade by β -adrenergic stimulation in rat ventricular myocytes. *J. Physiol.* **533**, 837–848 (2001).
85. Song, L. S. *et al.* β -Adrenergic stimulation synchronizes intracellular Ca^{2+} release during excitation-contraction coupling in cardiac myocytes. *Circ. Res.* **88**, 794–801 (2001).
86. Ginsburg, K. S. & Bers, D. M. Isoproterenol does not increase the intrinsic gain of cardiac excitation–contraction coupling (ECC). *Biophys. J.* **80**, 590a (2001).
87. Eisner, D. A., Choi, H. S., Diaz, M. E., O'Neill, S. C. & Trafford, A. W. Integrative analysis of calcium cycling in cardiac muscle. *Circ. Res.* **87**, 1087–1094 (2000).
88. Davare, M. A. *et al.* A β_2 adrenergic receptor signaling complex assembled with the Ca^{2+} channel $\text{Ca}_v1.2$. *Science* **293**, 298–301 (2001).
89. Bers, D. M. & Zioilo, M. T. When is cAMP not cAMP? Effects of compartmentalization. *Circ. Res.* **89**, 373–375 (2001).
90. Kuschel, M. *et al.* β_2 -adrenergic cAMP signaling is uncoupled from phosphorylation of cytoplasmic proteins in canine heart. *Circulation* **99**, 2458–2465 (1999).
91. Rybin, V. O., Xu, X. & Steinberg, S. F. Activated protein kinase C isoforms target to cardiomyocyte caveolae: stimulation of local protein phosphorylation. *Circ. Res.* **84**, 980–988 (1999).
92. Vila Petroff, M. G., Egan, J. M., Wang, X. & Sollott, S. J. Glucagon-like peptide-1 increases cAMP but fails to augment contraction in adult rat cardiac myocytes. *Circ. Res.* **89**, 445–452 (2001).
93. Aprigliano, O., Rybin, V. O., Pak, E., Robinson, R. B. & Steinberg, S. F. β_1 - and β_2 -adrenergic receptors exhibit differing susceptibility to muscarinic accentuated antagonism. *Am. J. Physiol.* **272**, H2726–H2735 (1997).

Seven-transmembrane-spanning receptors and heart function

Howard A. Rockman, Walter J. Koch* & Robert J. Lefkowitz†‡

Departments of Medicine and Cell Biology, *Surgery, †Biochemistry, and ‡The Howard Hughes Medical Institute, Duke University Medical Center, Durham, North Carolina 27710, USA (e-mail: h.rockman@duke.edu)

Understanding precisely how the heart can recognize and respond to many different extracellular signalling molecules, such as neurotransmitters, hormones and growth factors, will aid the identification of new therapeutic targets through which cardiovascular diseases can be combated. In recent years, we have learned more about the complex interactions that occur between the receptors and the signalling pathways of the heart and its environment. Most of these discoveries have focused on the most common type of cardiac receptor — the seven-transmembrane-spanning receptor or G-protein-coupled receptor.

The family of G-protein-coupled receptors (GPCRs) contain a conserved structure of seven transmembrane α -helices. Of these receptors, the adrenergic receptors and muscarinic cholinergic receptors are particularly important for the heart because they function in the homeostatic regulation of the cardiovascular system. The adrenergic receptors have been well studied and are the targets for many pharmaceuticals, such as the α - and β -blockers that are used widely to treat hypertension and myocardial diseases¹.

At least nine subtypes of adrenergic receptors have been cloned, including three α_1 -, three α_2 - and three β -adrenergic receptors (ARs)¹. In the heart, the β_1 -AR is the most predominant subtype, comprising 75–80% of total β -ARs; in contrast, there are only a small number of α -ARs, with a ratio of β - to α -ARs of about 10:1 in the human myocardium (ref. 1; and Table 1). An example of homeostatic regulation involving the normal heart is the control of heart rate. Resting heart rate is controlled by cholinergic signals mediated through muscarinic G_{α_i} -coupled receptors (where i indicates the inhibitory G protein), both through its negative action on adenylyl cyclase and through inhibition of the sinoatrial node by activation through the effects of the G_{α_i} and $G\beta\gamma$ subunits². During exercise, the withdrawal of parasympathetic influences and the increase in noradrenaline release from sympathetic postganglionic nerve endings increase cardiac output by increasing heart rate and enhancing contractility. These last effects are the result of enhanced β -AR signalling in myocytes.

Classical models have explained the regulation of cardiac function through second messenger systems operating by

means of the potentiation of cellular processes such as excitation–contraction coupling, contractile and relaxation processes, and heart rate. But emerging concepts are providing a glimpse of the complexity of this system; for example, it is now appreciated that adrenergic receptors do not simply function to generate second messengers, but rather activate a host of signalling proteins and pathways that are vitally important in the control of cardiac function, myocyte growth and programmed cell death (apoptosis). The discovery of these new GPCR signalling pathways and their application as therapeutic modalities are areas of intense investigation and the subjects of this review.

Traditional concepts of receptor regulation

When bound to their ligands, GPCRs are stabilized in an active conformation and stimulate heterotrimeric guanine-nucleotide-binding regulatory proteins (G proteins) through their intracellular domains. GPCRs lack catalytic activity, but interaction of the receptor with an agonist promotes the dissociation of G proteins into $G\alpha$ and $G\beta\gamma$ subunits. The G protein subunits (both $G\alpha$ and $G\beta\gamma$) then amplify and propagate signals inside the cell by modulating the activity of one or more effector molecules, including adenylyl cyclases, phospholipases and ion channels². In turn, the activity of these effector enzymes and ion channels regulates the production of second messenger molecules, which elicit cellular responses by activating different signalling pathways (Fig. 1). Examples of second messengers include cyclic AMP produced by stimulating β -ARs, and inositol-1,4,5-trisphosphate (InsP_3) and diacylglycerol (DAG) produced by stimulating α_1 -ARs.

Table 1 Important seven-transmembrane-spanning receptors

Receptor	β_1	β_2	β_3	$\alpha_{1A/B/D}/AT_1/ET$	α_2	M_2
Primary G protein	G_s	G_s/G_i	G_s/G_i	G_q/G_{11}	G_i	G_i
Tissue distribution	Heart	Heart, lung, vessels, kidney	Adipose, heart	Heart, vessels, smooth muscle	Coronary vessels, CNS, pancreas, platelets	Heart
Primary effector in heart tissue	AC, L-type Ca^{2+} channel	AC, L-type Ca^{2+} channel	AC	PLC- β	AC	AC, K^+ channels
Signals	$\uparrow\text{cAMP/PKA}$	$\uparrow\text{cAMP/PKA}$, MAPK	$\uparrow\text{cAMP/PKA}$	$\uparrow\text{DAG/InsP}_3$, PKC, MAPK	$\downarrow\text{cAMP/PKA}$	$\downarrow\text{cAMP/PKA}$, $\uparrow\text{PI}(3\text{K})$
Endogenous agonist	NA, A	NA, A	NA, A	NA, A, angiotensin II, endothelin	NA, A	ACh

α , α -adrenergic receptor subtypes A, B, D; β , β -adrenergic receptor subtypes 1, 2, 3; AC, adenylyl cyclase; ACh, acetylcholine; A, adrenaline; AT_1 , angiotensin-II receptor subtype 1; cAMP, adenosine 3',5' monophosphate; DAG, diacylglycerol; ET, endothelin receptor; InsP_3 , inositol-1,4,5-trisphosphate; M_2 , muscarinic cholinergic receptor subtype 2; MAPK, mitogen-activated protein kinase; NA, noradrenaline; PKA, protein kinase A; PKC, protein kinase C; PLC- β , phospholipase C- β . Adapted from ref. 1.

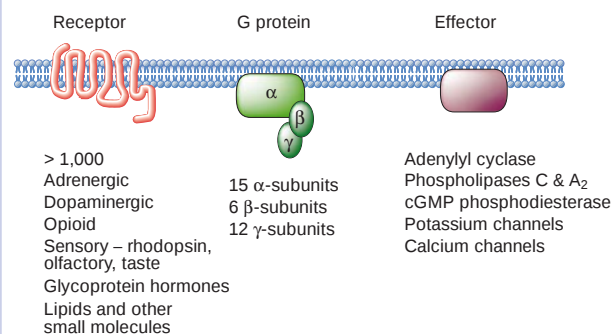


Figure 1 Summary of G-protein-coupled receptor signalling. The basic unit of G-protein-coupled receptor signalling comprises three parts: a seven-transmembrane-spanning receptor; a heterotrimeric G protein (figures next to each subunit indicate the number of variants that have been identified); and an effector, such as an enzyme or ion channel.

Different G-protein complexes are involved in transmitting signals to the different effector molecules that exist on cells. For the effector adenylyl cyclase, G_{α_s} is stimulatory as a consequence of β -AR activation, whereas G_{α_i} is inhibitory after muscarinic receptor stimulation. The signalling molecules $InsP_3$ and DAG are generated by the effector phospholipase C, which is activated by the G protein G_{α_q} in response to stimulation of the α_1 -AR, the angiotensin II receptor, or the endothelin receptor. Activation of these GPCRs is essential to the stress responses of the heart³.

An important adaptive mechanism of GPCRs is 'desensitization' — that is, the waning of receptor responsiveness despite continuing agonist stimulation. GPCRs have developed elaborate means of turning off their signal⁴. One mechanism for desensitization (which operates over hours) is receptor 'downregulation'; this refers to the net loss of receptors from the cell owing to a decrease in receptor synthesis, a destabilization of receptor messenger RNA or an increase in receptor degradation^{4,5}. The signal can also be terminated by uncoupling the receptor from its signal-transducing G protein (that

is, G_{α_s}) through a rapid process (seconds to minutes) that requires receptor phosphorylation by protein kinases.

The second messengers cAMP and DAG activate, respectively, protein kinase A (PKA) and protein kinase C (PKC), which phosphorylate GPCRs, leading to a process known as heterologous desensitization or 'non-agonist-specific' desensitization. In contrast, a mechanism of 'agonist-specific', or homologous desensitization of β -ARs is mediated by members of the GPCR kinase (GRK) family⁵. The six family members (GRK1–6) share the characteristic of phosphorylating only agonist-occupied or stimulated GPCRs^{4,5}. Some of the GRKs also have trivial names such as rhodopsin kinase (GRK1) or β -adrenergic receptor kinase, β -ARK1 (GRK2). With the exception of GRK1 (found almost exclusively in the retina) and GRK4 (expressed only in testes and brain regions), GRKs are expressed in all tissues including the heart⁵.

GRK-mediated phosphorylation of agonist-occupied GPCRs enhances the affinity of the receptor for interaction with cytosolic proteins known as the β -arrestins, which bind to the receptor and interdict further G-protein coupling. For β -ARs, this process prevents further activation of G_{α_s} and the subsequent stimulation of adenylyl cyclase. Increased GRK activity forces the equilibrium of β -ARs towards the inactive state, creating a condition of reduced responsiveness to catecholamines^{4,5}.

Evolving concepts of receptor regulation

In addition to their now classical role in desensitizing or turning off GPCR-stimulated generation of second messengers, the GRKs and β -arrestins function in previously unsuspected ways to actually promote many aspects of heptahelical receptor regulation and signalling. Recently, two main types of roles have been elucidated relating to the internalization of the receptors⁶ and to their signalling by assembling complex cascades, such as various mitogen-activated protein kinase (MAPK) pathways (refs 7–9; and Fig. 2).

Internalization of GPCRs, also known as sequestration or endocytosis, is a general albeit not universal feature of the biology of these receptors. Such ligand-promoted internalization seems to subserve several functions, including resensitization (by dephosphorylating the receptors in acidic endosomes), as well as complex, poorly defined signalling processes. GPCRs can be internalized by many mechanisms, including classical clathrin-coated pit-mediated processes, caveolae, and also non-coated pit mechanisms¹⁰. β -Arrestin

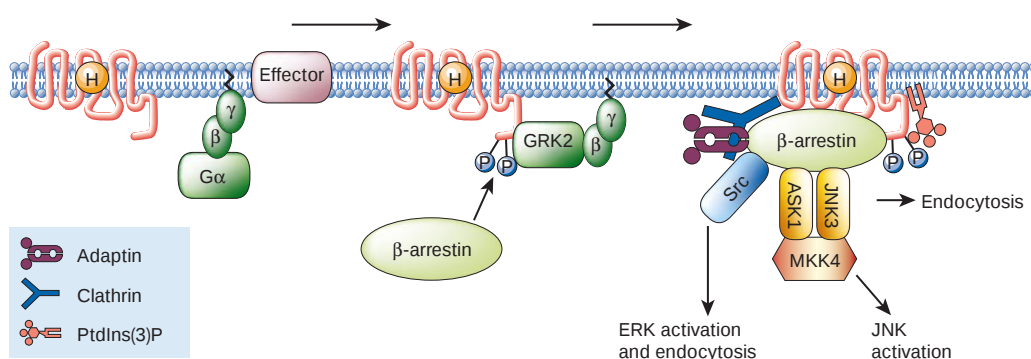


Figure 2 The many signalling roles of seven-transmembrane-spanning receptors. Agonist binding to the receptor results in the dissociation of heterotrimeric G proteins into G_{α} -GTP and $G_{\beta\gamma}$ subunits, which activate G-protein effectors. One consequence of $G_{\beta\gamma}$ subunit release is enhanced GRK-mediated phosphorylation of the agonist-occupied receptor. β -Arrestin binds to the GRK-phosphorylated receptor and not only uncouples the receptor from its G protein, but also acts as a scaffold to recruit clathrin and the clathrin adaptor AP2, which target the activated receptor for endocytosis. β -Arrestin-mediated interaction with Src can promote activation of

extracellular-regulated kinase (ERK) and endocytosis, by phosphorylating members of the internalization complex such as dynamin. The interaction of β -arrestin with a MAPK kinase (for example, ASK1) and c-Jun N-terminal kinase 3 (JNK3) can lead to the assembly of the JNK module with a MAPK kinase (for example, MKK4), which leads to activation of the MAPK JNK3. Thus the β -arrestins seem to act as adaptors that link GPCRs to both clathrin and the clathrin adaptor AP2, and as scaffolds that bring together other elements of a receptor complex to facilitate the activation of signalling pathways.

seems to be crucial in mediating the clathrin-coated pit mechanism of receptor internalization. Although the biology is not understood fully, β -arrestins seem to act as adaptors that link the receptors not only to clathrin, but also to the clathrin adaptor AP2, and possibly to other required elements of the internalization system⁶.

The second recently elucidated function of β -arrestins (Fig. 2) is their role as adaptors or scaffolds that bring together elements of complex heptahelical-receptor-stimulated signalling pathways, and that facilitate the activation and subcellular localization of the signalling proteins involved in these pathways^{7,11,12}. This function may be particularly important for understanding essential aspects of cardiac function. As noted above, heptahelical receptors, unlike receptor tyrosine kinases, do not contain any intrinsic catalytic activity. But many of their functions are mediated through the activation of non-receptor tyrosine kinases, such as members of the Src family.

β -Arrestins can act as adaptors that bind members of the Src family and, in a ligand-dependent fashion, can bring these tyrosine kinases into complex with heptahelical receptors⁷⁻⁹. Often, this is a means of linking the GPCRs with elements of the MAPK cascades such as the extracellular regulated kinase (ERK) cascade. Moreover, β -arrestins can act as a scaffold for the various members of several MAPK cascades (that is, MAPK kinase kinase, MAPK kinase and MAPK). This scaffolding function has been shown for both the c-Jun amino-terminal kinase 3 (JNK3)¹² and ERK^{9,11} modules, and brings their activation and subcellular localization under the direct control of the stimulated receptor. Such regulation has been documented most clearly for the angiotensin II type 1A receptor, protease-activated receptor 2 and the neurokinin receptor^{7,9,11}.

Another recently discovered function of β -arrestin involves ubiquitination¹³ — a process that marks proteins for degradation. Although agonist-dependent internalization, recycling and degradation of GPCRs have been studied extensively, few studies have linked ubiquitination with the trafficking of mammalian GPCRs. But now it seems that both β_2 -AR and β -arrestin undergo rapid agonist-stimulated ubiquitination to fulfil at least two roles: first, β -arrestin ubiquitination is necessary for receptor internalization; and second, receptor ubiquitination is necessary for lysosomal targeting and degradation of the receptors. Moreover, β -arrestin seems to act as an adapter that brings the ubiquitination machinery into the complex with the receptor¹³. These new data highlight the complexities in the regulation of GPCR function, and also remind us that our understanding of the area is continually evolving.

Adrenergic receptors and heart disease

Myocardial hypertrophy

Cardiac hypertrophy is the physiological response of the heart to an increased work load. Because adult myocardial cells are widely thought to be terminally differentiated, cardiac growth during the hypertrophic process results primarily from an increase in the size of individual myocardial cells with little or no change in cell number¹⁴. The biochemical features of the hypertrophic response are an increase in contractile protein content, myofilament organization, and re-expression of embryonic markers (such as atrial natriuretic factor, α -skeletal actin and β -myosin heavy chain), which is caused mainly by transcriptional activation of the genes that encode these proteins¹⁴. Regulation of myocyte cell growth by GPCRs involves a complex network of interacting pathways that activate not only key effector molecules such as *ras* and the MAPKs^{15,16}, but also phosphatidylinositol-3-OH kinase (PI(3)K)-dependent¹⁷ and calcineurin-dependent³ pathways. Importantly, cardiac hypertrophy is also characterized by marked abnormalities in β -AR function, which are partly caused by elevated myocardial concentrations of β -ARK1 (ref. 18).

In vitro studies have suggested that α_1 -ARs are pivotal in promoting cardiomyocyte hypertrophy¹⁹. In cardiac myocytes, G_q -coupled receptors (such as α_1 -ARs, angiotensin and endothelin receptors) can activate hypertrophic MAPK pathways²⁰. The MAPK superfamily

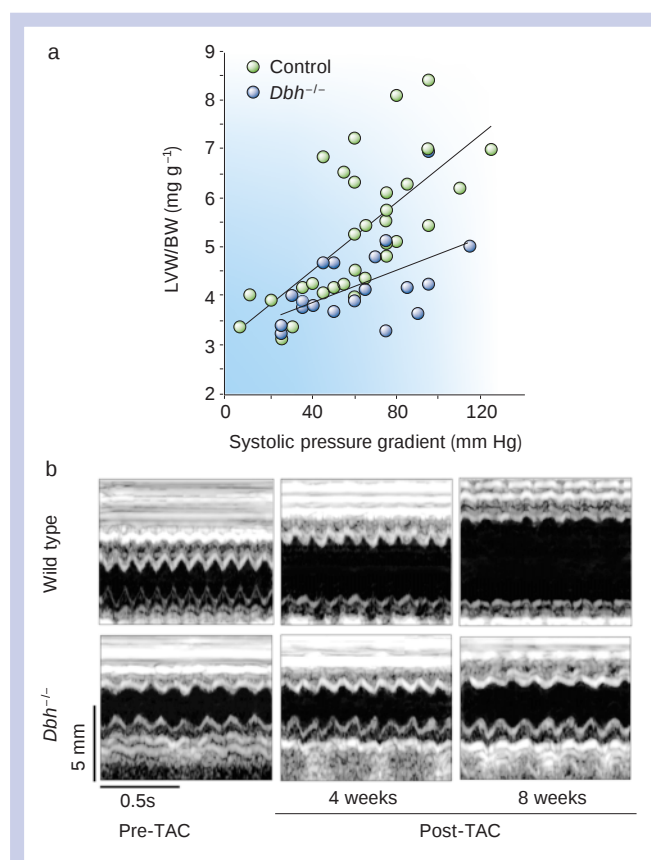


Figure 3 Attenuation of cardiac hypertrophy preserves cardiac function: reappraisal of current dogma. Pressure overload cardiac hypertrophy was induced by surgical constriction of the transverse aortic in genetically altered mice that lack endogenous noradrenaline and adrenaline, created by disruption of the dopamine β -hydroxylase gene (*Dbh*^{-/-}), the essential enzyme in the biosynthetic pathway converting dopamine to noradrenaline. **a**, Pressure overload *Dbh*^{-/-} mice develop less cardiac hypertrophy (shown as a reduction in left ventricular weight to body weight ratio, LVW/BW) compared with similarly pressure overloaded control mice. **b**, Representative serial M-mode echocardiographic tracings before, 4 weeks and 8 weeks after creation of mechanical pressure overload. Remarkably, even though cardiac hypertrophy in *Dbh*^{-/-} mice was significantly inhibited, there was no deterioration of cardiac function. Thus, blunting of the hypertrophic response, and the resultant increase in wall stress, is associated with less deterioration in cardiac function despite the continued presence of a pressure load on the heart. Adapted, with permission, from ref. 28.

regulates three principal pathways: the ERK pathway and two stress-activated protein kinase pathways mediated by JNKs and p38 MAPKs²⁰. The important role of GPCRs and MAPK signalling in the development of *in vivo* cardiac hypertrophy has been shown recently by the generation of transgenic mice overexpressing either G_{α_q} (ref. 21) or the angiotensin II AT₁ receptor²², and by the adenoviral-mediated transfer of a dominant inhibitory mutant of an upstream activator of JNK²³. To test whether G_q -coupled receptor signalling is required for the development of *in vivo* hypertrophy, recent experiments have used a transgenic mouse with cardiac-specific overexpression of a 54-amino-acid carboxy-terminal peptide of G_{α_q} . Overexpression of this peptide blocks signalling through G_q -coupled receptors, and in response to surgically induced *in vivo* pressure overload²⁴ both cardiac hypertrophy and the activation of MAPK are attenuated significantly^{15,25}.

It has been generally accepted that myocardial hypertrophy develops in response to increased cardiac workload as a compensatory mechanism that acts to normalize wall stress, thereby preventing the development of heart failure^{14,26}. Although this remodelling may indeed normalize wall stress, results from the Framingham Heart

Study show that there is an association between ventricular hypertrophy and increased cardiac mortality, which casts doubt on the validity of the 'wall stress' hypothesis²⁷. To determine whether cardiac hypertrophy is adaptive or maladaptive, two genetically altered mouse models with an attenuated hypertrophic response to pressure overload have been studied after long-term mechanical overload. These models were transgenic mice that overexpress the $G\alpha_i$ inhibitor peptide, and knock-out mice that lack endogenous noradrenaline and adrenaline²⁸. Remarkably, despite the lack of adequate hypertrophy to normalize wall stress in the genetically altered mice, there was less deterioration in heart function than in control mice with a 'normal' hypertrophic response (ref. 28; and Fig. 3).

Taken together, the above experiments show that G_q and catecholamine (that is, adrenaline and noradrenaline) signalling have a crucial role in activating the hypertrophic process. Moreover, contrary to previous hypotheses, hypertrophy itself and not elevated wall stress seems to be a trigger for cardiac decompensation. These data suggest that therapeutic agents that target the GPCR- G_q interface will be useful in counteracting the pathophysiological signalling that occurs in pressure overload.

Heart failure and β -adrenergic receptor signalling

Heart failure is characterized by left ventricular dysfunction associated with a complex of symptoms that relate to inadequate perfusion of tissues and pulmonary congestion (see review in this issue by Towbin and Bowles, pages 227–233). When the heart is damaged by an insult, activation of the sympathetic nervous system has an important role in adapting circulatory homeostasis to changes in environment²⁹. But chronic activation of sympathetic activity has adverse biological effects and can initiate or accelerate cardiovascular pathology³⁰. These observations have led to the hypothesis that sympathetic activation may be central to the progression of heart failure³¹ and that pharmacological interference with this system can produce haemodynamic and clinical benefits³².

In the failing human heart there are abnormalities at several levels of the β -AR pathways (Table 2). Chronic increases in catecholamines leads to both β -AR desensitization and downregulation³¹. Indeed, β_1 -ARs have been shown to be downregulated selectively, but both β_1 - and β_2 -ARs are markedly uncoupled from G proteins in failing hearts³¹—a process attributed to increased levels of myocardial β -ARK1 (ref. 33) and myocardial $G\alpha_i$ (ref. 34). This not only results in the attenuation of β -AR signalling, but may also enhance the coupling of receptors to other signalling pathways that are involved in ventricular remodelling, such as the MAPK and PI(3)K cascades^{35,36}.

Genetic heterogeneity in the structure of both the β_1 - and the β_2 -AR in the human population has been identified^{37,38}. β_2 -ARs containing a threonine to isoleucine substitution at amino acid 164 in the fourth transmembrane-spanning domain display decreased basal and agonist-stimulated adenylyl cyclase activities owing to defective coupling of the receptor to G_s (ref. 37). Individuals with heart failure that possess the Ile 164 allele have a significantly reduced survival³⁹ and a depressed capacity for exercise⁴⁰. These data show that genetic polymorphisms in β -ARs may lead to inter-individual

differences in the pathophysiological characteristics of people with chronic heart failure and will probably become an important part of medical practice by identifying high risk individuals⁴¹. For example, identifying individuals at high risk for sudden cardiac death would potentially lead to the implementation of more aggressive management such as early cardiac transplantation or the insertion of an implantable defibrillator pacemaker.

Heart failure and the effect of β -blockade

Because β -AR blockers are known to depress cardiac function, historically the use of these agents was considered to be contra-indicated in individuals with compromised ventricular function owing to their short-term adverse effects⁴². Recent data have now shown that, in contrast to the adverse short-term changes in ventricular function, the long-term biological actions of β -ARs blockers are beneficial in most cases. The most dramatic finding has been the effect of β -blockade on mortality. Several studies have been published that show remarkable improvement in the survival of individuals with moderate^{43–45} and severe⁴⁶ heart failure. These findings show how drugs interfering with the function of the β -AR system can be useful for treating heart failure and also that interfering with the chronic activation of the sympathetic system in individuals with impaired cardiac function can lead to beneficial outcomes.

Along with the effects on left ventricular contractile function, structural changes involving the size and shape of the heart may also account for the long-term benefits of treating heart failure with β -AR blockers. Treatment with the β -blocker carvedilol has been associated with a reduction in end-diastolic and end-systolic left ventricular diameters⁴⁷, supporting the hypothesis that β -blockade has a beneficial effect on the structure of the failing heart. It will be important to gain a broader understanding of the exact molecular mechanisms involved in these changes in the failing heart evoked by β -AR blocker treatment.

Mouse models to study the β -AR system

Manipulating receptor expression

Despite numerous investigations that describe the changes in the β -AR system during heart failure, it remains unknown whether chronic β -AR desensitization in heart failure is helpful or deleterious. One approach to resolving this issue is through the use of gene-targeted mice, which allows the expression of various components of the myocardial β -AR system to be manipulated. Combining this technology with sophisticated physiological measurements of cardiac haemodynamics provides a powerful set of tools to study the regulation of myocardial contractility⁴⁸.

Transgenic technology coupled with the use of cardiac-specific promoters has made it possible to target β -ARs and other components of the β -AR system specifically to the myocardium⁴⁹. Transgenic mice created with ventricular overexpression of either the β_1 - or the β_2 -AR show significantly different phenotypes, indicating a fundamental difference in their signalling and function in myocardium. A 5- to 15-fold overexpression of β_1 -ARs leads to a phenotype of dilated cardiomyopathy even in young mice⁵⁰, and is consistent with the pathology caused by chronic administration of catecholamines. In contrast, overexpression of up to 200 times more β_2 -ARs than the endogenous receptor does not cause any significant cardiac pathology in mice, which are instead characterized by enhanced biochemical and *in vivo* cardiac function^{51,52}. However, transgenic overexpression of very high levels of β_2 -ARs (> 200-fold) results in a rapid and progressive cardiomyopathic phenotype, consistent with the pathological toxicity caused by excess β -AR stimulation⁵².

The role of β -ARs in cardiac function has been addressed further by gene disruption experiments. Knockout of the β_1 -AR in mice generally leads to embryonic lethality, with most mice homozygous for the β_1 -AR gene disruption dying *in utero*⁵³. Embryonic mortality is reduced if the mice are outbred. The surviving mice have normal

Table 2 Alterations in the β -AR signalling pathway in chronic heart failure

Molecule	Change
β_1 -ARs	↓, uncoupled
β_2 -ARs	NC, uncoupled
β -ARK1	↑mRNA levels, ↑activity
GRK3	NC
GRK5	Not studied
β -Arrestin 1 and 2	NC
$G\alpha_i$	↑
$G\alpha_s$	NC

NC, No change; ↑, increased; ↓, decreased; data are from refs 31,33,34.

resting heart rate and blood pressures. Notably, despite the presence of β_2 -ARs in these mouse hearts, there is no contractility response to β -agonist stimulation, suggesting that β_1 -AR is the predominant mediator of catecholamine-induced chronotropy and inotropy in the mouse⁵³. In contrast, β_2 -AR knockout mice are viable and healthy; physiological consequences are observed only during the stress of exercise and are a result of alterations in both vascular tone and energy metabolism⁵⁴, suggesting that a lack of β_2 -ARs has little impact on cardiac function.

The stimulation of β_1 -ARs, as compared with β_2 -ARs, seems to have distinctly different effects in cardiac myocytes, and may be an explanation for the marked phenotypic difference observed when these receptors are overexpressed in the hearts of transgenic mice. This may be partly due to the ability of cardiac β_2 -ARs to couple to and activate both G_{α_s} and G_{α_i} proteins, whereas cardiac β_1 -ARs couple only to G_{α_s} . Dual coupling to G_{α_s} and G_{α_i} might provide a mechanism through which cardiac myocyte β_2 -ARs can elicit survival signals on agonist stimulation; in contrast, stimulating cardiac β_1 -ARs activates only apoptotic pathways^{36,55}. Although we can only speculate at this time, it is possible that the fundamentally different mechanisms by which β_1 - and β_2 -ARs signal in the myocardium may lead to the unique localization and recruitment of molecules, adaptors or scaffolds that promote both stimulatory and protective pathways rather than pure stimulatory pathways.

Notably, the targeted overexpression of β_2 -AR in the hearts of experimental models of heart failure may also be therapeutic. Whereas moderate (~60-fold) overexpression of cardiac β_2 -AR in a mouse model prevented hypertrophy and improved cardiac function⁵⁶, marked overexpression (> 100-fold) caused a worsening of the disease^{57,58}. The latter observation probably represents the known pathological effect of very high levels of overexpression of the β_2 -AR; however, it seems that modest levels of β_2 -AR expression can improve heart function. Indeed, in rabbit models, recombinant adenoviruses encoding β_2 -AR delivered by various strategies have brought about a significant improvement in myocardial β -AR signalling and enhanced *in vivo* cardiac function^{59–61}.

Other approaches have also rescued mouse models of heart failure, such as mice deficient in phospholamban⁶² — the sarcoplasmic reticulum Ca^{2+} -ATPase inhibitor — and transgenic overexpression of adenylyl cyclase type VI (ref. 63). These studies highlight further the great potential of using mouse models to identify and test potential therapeutic targets.

Manipulating β -ARK1 activity

Myocardial regulation of β -ARs *in vivo* has been studied through genetic manipulation of GRKs in the mouse^{64–67}. Using different strategies, the concentration and activity of β -ARK1 — the primary GRK expressed in the heart — has been manipulated in the hearts of mice by overexpressing the full-length protein, by overexpressing a peptide inhibitor of β -ARK (β -ARKct)⁶⁴ or by ablating the β -ARK1 gene⁶⁵. The 194-amino-acid β -ARKct peptide is particularly interesting because it contains the $G\beta\gamma$ -binding domain of β -ARK1, which competes with endogenous β -ARK1 for $G\beta\gamma$ -mediated membrane translocation and activation. From studies that have carefully characterized the cardiac phenotype of these genetically altered mice, we have learned that, in the normal heart *in vivo*, the level of contractile function and the heart's response to catecholamine stimulation is determined primarily by the level of β -ARK1 activity^{64–67}. These data have important implications for disease states characterized by increased β -ARK1 activity, because even partial inhibition of β -ARK1 activity may lead to improved functional responsiveness to catecholamines (see below).

The growing number of genetic mouse models of cardiomyopathy and heart failure has been a great advantage to cardiovascular investigators. Generally these extremely valuable strains of mice are discovered serendipitously when mice in which a particular gene of

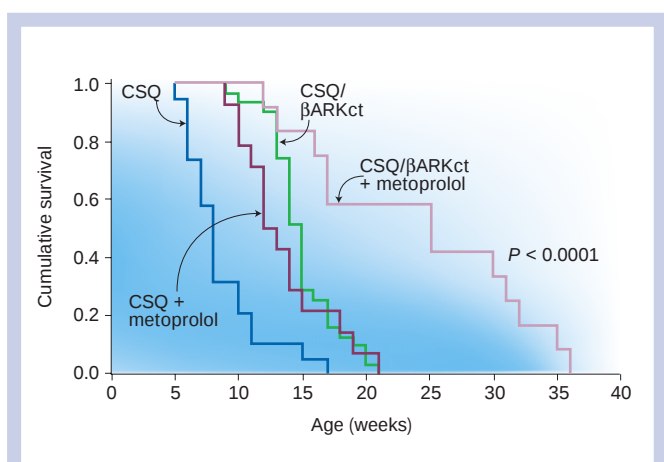


Figure 4 Improved survival of transgenic mice overexpressing calsequestrin (CSQ) through inhibition of β -ARK by overexpressing the β -ARKct peptide. Kaplan–Meier analysis of the survival probability between the specified genotypes of mice either untreated or chronically treated with the selective β_1 -AR antagonist metoprolol. Mean survival age of the CSQ/ β -ARKct mice is significantly higher than that of CSQ transgenic littermates ($P < 0.0001$). Moreover, treating the CSQ/ β -ARKct mice with metoprolol improves survival further, indicating that there is synergism between the two therapeutic modalities. Adapted, with permission, from ref. 71.

interest has been either knocked out or overexpressed are found to develop features of cardiomyopathy and heart failure. In several cases these mouse models closely recapitulate many of the features of human heart failure^{21,68–70}. To test whether β -ARK1 has an important role in the pathophysiology of cardiac failure, transgenic mice with cardiac overexpression of β -ARKct (which results in β -ARK inhibition) have been mated with several of these mouse models^{48,57,58}. These have included the muscle *lim* protein (MLP) knockout⁵⁷, cardiac overexpression of calsequestrin (ref. 71; and Fig. 4), and transgenic overexpression of a mutant cardiac α -myosin heavy chain⁷⁰. Strikingly, in each of these cases cardiac overexpression of β -ARKct resulted in prevention of progressive deterioration in cardiac dysfunction, prevention of hypertrophy, improved exercise tolerance, correction in the classical biochemical alterations of β -AR function³¹ and, where it could be assessed, markedly improved survival. Moreover, effects on survival were synergistic with those achieved by β -AR blocker treatment⁷¹ — an interesting finding given recent clinical data that show improved survival of individuals with chronic heart failure that were treated with β -blockers^{43–46}.

Additional studies have documented the ability of replication-deficient adenoviruses encoding the β -ARKct transgene to positively influence cardiac function in larger animal models. The β -ARKct adenoviral transgene, delivered by different strategies to a rabbit model of experimental heart failure, prevented biochemical abnormalities in the β -AR system, delayed the onset of haemodynamic heart failure and reversed the development of heart failure postmyocardial infarction^{59,72,73}. Taken together, overwhelming evidence supports the hypothesis that the loss of β -AR signalling — the *sine qua non* of the human heart failure phenotype — contributes to the pathogenesis of the syndrome. Furthermore, these data suggest that genetic manipulation and restoration of adrenergic receptor signalling represent a new therapeutic approach to treatment of the failing heart.

Although the above studies show the salutary effects of β -ARK1 inhibition, the strategy used to block action on β -ARK1 *in vivo* has been to overexpress the β -ARKct peptide. It should be borne in mind that β -ARKct also functions more generally to sequester $G\beta\gamma$ subunits from other biologically important interactions, such as those involved in the activation of $PI(3)K$ ^{17,74} and $I_{K,ACh}$

channels². The extent to which, if any, interruption of such interactions might contribute to the salutary effects of β -ARKct remains to be determined.

The paradox of β -AR blockade and β -ARK1 inhibition

An important issue raised by the above findings is how to rationalize the seemingly protective effect of interrupting β -AR signalling with β -AR blockers and the beneficial effect of inhibiting β -ARK1 in heart failure. In fact, inhibiting β -ARK1 activity would be expected to facilitate β -AR signalling — much like β -agonists, which are poorly tolerated when administered chronically to individuals with heart failure⁷⁵. In this context, we need to address several points that may aid our understanding of this apparent paradox.

First, at a molecular level β -ARK1 inhibition actually shares many attributes with β -AR blockade rather than with β -agonism. For example, β -agonists promote further desensitization and receptor downregulation through constant bombardment of the β -AR system, whereas β -ARK1 inhibition may return β -ARs to a more normal state of signalling because desensitization will be inhibited and the system will only be facilitated during the normal flow of catecholamines. Second, as discussed above, emerging models now suggest that β -ARK1-mediated phosphorylation of β -AR promotes β -AR internalization, which may be associated with the activation of signalling pathways involved in cell growth and survival. Limiting the activation of these pathways by inhibiting β -ARK1 may limit the growth response of the myocyte. Indeed, the growth response may itself be maladaptive.

Third, the fact that the β -AR system is severely crippled in the failing heart no doubt perpetuates the hyperactivity of the sympathetic nervous system, which can initiate or accelerate cardiovascular pathology and provoke clinical events in the presence of cardiovascular disease. By relieving a brake on the system (that is, by β -ARK1 inhibition), cardiac function can be improved, and this may lead to a dampening of sympathetic overdrive. Fourth, chronic β -AR blockade also 'normalizes' the myocardial β -AR system primarily through receptor upregulation (or at least by preventing further downregulation) so that when catecholamine stimulation is needed to enhance cardiac performance acutely, such as during exercise and stress, signalling through β -ARs is actually facilitated. In fact, it is possible that the beneficial effects of β -blockers in heart failure involve, at least in part, regulation of β -ARK1, as studies have shown that chronic β -AR blockade decreases the expression of β -ARK1 in the heart and reduces cardiac activity of GRK⁷⁶.

Fifth, overexpression of the β -ARKct peptide restores normal responsiveness to β -agonist stimulation^{48,57}, which suggests that abnormal β -AR-G-protein coupling, early in the development of the failing heart, may be a crucial event in the progressive deterioration of cardiac pump function. Last, it should be noted that the action of β -ARK1 in the heart is not limited to β -ARs.

The hypothesis that β -ARK1 inhibition has little similarity to chronic β -AR agonism is supported further by findings in the calsequestrin-overexpressing mouse model of heart failure. The combination of β -ARKct expression and chronic metoprolol (a β_1 -AR selective antagonist) treatment has been shown to increase survival and improve cardiac function in these animals, suggesting that these two therapies may act synergistically to reduce lethal arrhythmias and promote favourable remodelling⁷¹.

Building on new receptor knowledge

As clinical trials have shown for the use of β -AR blockers in individuals with heart failure, scientific data that were previously considered to be dogma can later prove not to be valid. (For example, treatment with β -AR blockers was contra-indicated because of their strong negative inotropic effects, and that normalization of wall stress was needed to preserve cardiac function with pressure overload.) The identification of new GPCR-interacting proteins and signalling pathways will hopefully lead to the development of small molecules

that model the action of these new targets (for example, the β -ARKct), and the development of new transgenic animal models will provide an opportunity to test these molecules in several experimental and clinical conditions of disease (for example, endothelin receptor antagonists, which are currently in clinical trial for heart failure). A glimpse of what lies ahead for the field of cardiovascular disease has been shown recently for individuals with chronic heart failure treated with β -blockers⁷⁷. □

- Hoffman, B. B. & Lefkowitz, R. J. in *Goodman and Gilman's The Pharmacological Basis of Therapeutics* 9th edn (eds Hardman, J. G., Gilman, A. G. & Limbird, L. E.) 199–248 (McGraw-Hill, New York, 1996).
- Clapham, D. E. & Neer, E. J. G protein $\beta\gamma$ subunits. *Annu. Rev. Pharmacol. Toxicol.* **37**, 167–203 (1997).
- Molkentin, J. D. & Dorn, I. G. II Cytoplasmic signaling pathways that regulate cardiac hypertrophy. *Annu. Rev. Physiol.* **63**, 391–426 (2001).
- Lefkowitz, R. J. G protein-coupled receptors. III. New roles for receptor kinases and β -arrestins in receptor signaling and desensitization. *J. Biol. Chem.* **273**, 18677–18680 (1998).
- Pitcher, J. A., Freedman, N. J. & Lefkowitz, R. J. G protein-coupled receptor kinases. *Annu. Rev. Biochem.* **67**, 653–692 (1998).
- Laporte, S. A., Oakley, R. H., Holt, J. A., Barak, L. S. & Caron, M. G. The interaction of β -arrestin with the AP-2 adaptor is required for the clustering of β_2 -adrenergic receptor into clathrin-coated pits. *J. Biol. Chem.* **275**, 23120–23126 (2000).
- DeFea, K. A. et al. The proliferative and antiapoptotic effects of substance P are facilitated by formation of a β -arrestin-dependent scaffolding complex. *Proc. Natl Acad. Sci. USA* **97**, 11086–11091 (2000).
- Luttrell, L. M. et al. β -arrestin-dependent formation of β_2 adrenergic receptor–Src protein kinase complexes. *Science* **283**, 655–661 (1999).
- DeFea, K. A. et al. β -arrestin-dependent endocytosis of proteinase-activated receptor 2 is required for intracellular targeting of activated ERK1/2. *J. Cell Biol.* **148**, 1267–1281 (2000).
- Schmid, S. L. Clathrin-coated vesicle formation and protein sorting: an integrated process. *Annu. Rev. Biochem.* **66**, 511–548 (1997).
- Luttrell, L. M. et al. Activation and targeting of extracellular signal-regulated kinases by β -arrestin scaffolds. *Proc. Natl Acad. Sci. USA* **98**, 2449–2454 (2001).
- McDonald, P. H. et al. β -arrestin 2: a receptor-regulated MAPK scaffold for the activation of JNK3. *Science* **290**, 1574–1577 (2000).
- Shenoy, S. K., McDonald, P. H., Kohout, T. A. & Lefkowitz, R. J. Regulation of receptor fate by ubiquitination of activated β_2 -adrenergic receptor and β -arrestin. *Science* **294**, 1307–1313 (2001).
- Chien, K. R. Stress pathways and heart failure. *Cell* **98**, 555–558 (1999).
- Esposito, G. et al. Cardiac overexpression of a G_i inhibitor blocks induction of extracellular signal-regulated kinase and c-Jun NH₂-terminal kinase activity in *in vivo* pressure overload. *Circulation* **103**, 1453–1458 (2001).
- Rapacciuolo, A. et al. Important role of endogenous norepinephrine and epinephrine in the development of *in vivo* pressure-overload cardiac hypertrophy. *J. Am. Coll. Cardiol.* **38**, 876–882 (2001).
- Naga Prasad, S. V., Esposito, G., Mao, L., Koch, W. J. & Rockman, H. A. $G\beta\gamma$ -dependent phosphoinositide 3-kinase activation in hearts with *in vivo* pressure overload hypertrophy. *J. Biol. Chem.* **275**, 4693–4698 (2000).
- Choi, D. J., Koch, W. J., Hunter, J. J. & Rockman, H. A. Mechanism of β -adrenergic receptor desensitization in cardiac hypertrophy is increased β -adrenergic receptor kinase. *J. Biol. Chem.* **272**, 17223–17229 (1997).
- Knowlton, K. U. et al. The α 1A-adrenergic receptor subtype mediates biochemical, molecular, and morphologic features of cultured myocardial cell hypertrophy. *J. Biol. Chem.* **268**, 15374–15380 (1993).
- Sugden, P. H. Signaling in myocardial hypertrophy: life after calcineurin? *Circ. Res.* **84**, 633–646 (1999).
- D'Angelo, D. D. et al. Transgenic G_{α_i} overexpression induces cardiac contractile failure in mice. *Proc. Natl Acad. Sci. USA* **94**, 8121–8126 (1997).
- Hein, L. et al. Overexpression of angiotensin AT₁ receptor transgene in the mouse myocardium produces a lethal phenotype associated with myocyte hyperplasia and heart block. *Proc. Natl Acad. Sci. USA* **94**, 6391–6396 (1997).
- Choukroun, G. et al. Regulation of cardiac hypertrophy *in vivo* by the stress-activated protein kinases/c-Jun NH₂-terminal kinases. *J. Clin. Invest.* **104**, 391–398 (1999).
- Rockman, H. A. et al. Segregation of atrial-specific and inducible expression of an atrial natriuretic factor transgene in an *in vivo* murine model of cardiac hypertrophy. *Proc. Natl Acad. Sci. USA* **88**, 8277–8281 (1991).
- Akhter, S. A. et al. Targeting the receptor– G_i interface to inhibit *in vivo* pressure overload myocardial hypertrophy. *Science* **280**, 574–577 (1998).
- Grossman, W., Jones, D. & McLaurin, L. P. Wall stress and patterns of hypertrophy in the human left ventricle. *J. Clin. Invest.* **56**, 56–64 (1975).
- Levy, D., Garrison, R. J., Savage, D. D., Kannel, W. B. & Castelli, W. P. Prognostic implications of echocardiographically determined left ventricular mass in the Framingham heart study. *N. Engl. J. Med.* **322**, 1561–1566 (1990).
- Esposito, G. et al. Genetic alterations that inhibit *in vivo* pressure overload hypertrophy prevent cardiac dysfunction despite increased wall stress. *Circulation* **105**, 85–92 (2002).
- Leimbach, W. N. Jr et al. Direct evidence from intraneural recordings for increased central sympathetic outflow in patients with heart failure. *Circulation* **73**, 913–919 (1986).
- Cohn, J. N. et al. Plasma norepinephrine as a guide to prognosis in patients with chronic congestive heart failure. *N. Engl. J. Med.* **311**, 819–823 (1984).
- Bristow, M. R. Why does the myocardium fail? Insights from basic science. *Lancet* **352** (Suppl. 1), 8–14 (1998).
- Packer, M. The neurohormonal hypothesis: a theory to explain the mechanism of disease progression in heart failure. *J. Am. Coll. Cardiol.* **20**, 248–254 (1992).
- Ungerer, M. et al. Expression of β -arrestins and β -adrenergic receptor kinases in the failing human heart. *Circ. Res.* **74**, 206–213 (1994).
- Feldman, A. M. et al. Increase of the 40,000-mol wt pertussis toxin substrate (G protein) in the failing human heart. *J. Clin. Invest.* **82**, 189–197 (1988).
- Daaka, Y., Luttrell, L. M. & Lefkowitz, R. J. Switching of the coupling of the β_2 -adrenergic receptor to different G proteins by protein kinase A. *Nature* **390**, 88–91 (1997).

36. Zhu, W. Z. *et al.* Dual modulation of cell survival and cell death by β_2 -adrenergic signaling in adult mouse cardiac myocytes. *Proc. Natl Acad. Sci. USA* **98**, 1607–1612 (2001).
37. Green, S. A., Cole, G., Jacinto, M., Innis, M. & Liggett, S. B. A polymorphism of the human β_2 -adrenergic receptor within the fourth transmembrane domain alters ligand binding and functional properties of the receptor. *J. Biol. Chem.* **268**, 23116–23121 (1993).
38. Podlowski, S. *et al.* β_1 -adrenoceptor gene variations: a role in idiopathic dilated cardiomyopathy? *J. Mol. Med.* **78**, 87–93 (2000).
39. Liggett, S. B. *et al.* The Ile164 β_2 -adrenergic receptor polymorphism adversely affects the outcome of congestive heart failure. *J. Clin. Invest.* **102**, 1534–1539 (1998).
40. Wagoner, L. E. *et al.* Polymorphisms of the β_2 -adrenergic receptor determine exercise capacity in patients with heart failure. *Circ. Res.* **86**, 834–840 (2000).
41. Collins, F. S. Shattuck lecture—medical and societal consequences of the human genome project. *N. Engl. J. Med.* **341**, 28–37 (1999).
42. Epstein, S. E. & Braunwald, E. The effect of β -adrenergic blockade on patterns of urinary sodium excretion. Studies in normal subjects and in patients with heart disease. *Ann. Intern. Med.* **65**, 20–27 (1966).
43. The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): a randomised trial. *Lancet* **353**, 9–13 (1999).
44. Effect of metoprolol CR/XL in chronic heart failure: metoprolol CR/XL randomised intervention trial in congestive heart failure (MERIT-HF). *Lancet* **353**, 2001–2007 (1999).
45. Packer, M. *et al.* The effect of carvedilol on morbidity and mortality in patients with chronic heart failure. *N. Engl. J. Med.* **334**, 1349–1355 (1996).
46. Packer, M. *et al.* Effect of carvedilol on survival in severe chronic heart failure. *N. Engl. J. Med.* **344**, 1651–1658 (2001).
47. Australia/New Zealand Heart Failure Research Collaborative Group. Randomised, placebo-controlled trial of carvedilol in patients with congestive heart failure due to ischaemic heart disease. *Lancet* **349**, 375–380 (1997).
48. Esposito, G. *et al.* Cellular and functional defects in a mouse model of heart failure. *Am. J. Physiol. Heart Circ. Physiol.* **279**, H3101–H3112 (2000).
49. Koch, W. J., Lefkowitz, R. J. & Rockman, H. A. Functional consequences of altering myocardial adrenergic receptor signaling. *Annu. Rev. Physiol.* **62**, 237–260 (2000).
50. Engelhardt, S., Hein, L., Wiesmann, F. & Lohse, M. J. Progressive hypertrophy and heart failure in β_1 -adrenergic receptor transgenic mice. *Proc. Natl Acad. Sci. USA* **96**, 7059–7064 (1999).
51. Milano, C. A. *et al.* Enhanced myocardial function in transgenic mice overexpressing the β_2 -adrenergic receptor. *Science* **264**, 582–586 (1994).
52. Liggett, S. B. *et al.* Early and delayed consequences of β_2 -adrenergic receptor overexpression in mouse hearts: critical role for expression level. *Circulation* **101**, 1707–1714 (2000).
53. Rohrer, D. K. Physiological consequences of β -adrenergic receptor disruption. *J. Mol. Med.* **76**, 764–772 (1998).
54. Chruscinski, A. J. *et al.* Targeted disruption of the β_2 adrenergic receptor gene. *J. Biol. Chem.* **274**, 16694–16700 (1999).
55. Communal, C., Singh, K., Sawyer, D. B. & Colucci, W. S. Opposing effects of β_1 - and β_2 -adrenergic receptors on cardiac myocyte apoptosis: role of a pertussis toxin-sensitive G protein. *Circulation* **100**, 2210–2212 (1999).
56. Dorn, G. W. II, Tepe, N. M., Lorenz, J. N., Koch, W. J. & Liggett, S. B. Low- and high-level transgenic expression of β_2 -adrenergic receptors differentially affect cardiac hypertrophy and function in $G\alpha_q$ -overexpressing mice. *Proc. Natl Acad. Sci. USA* **96**, 6400–6405 (1999).
57. Rockman, H. A. *et al.* Expression of a β -adrenergic receptor kinase 1 inhibitor prevents the development of myocardial failure in gene-targeted mice. *Proc. Natl Acad. Sci. USA* **95**, 7000–7005 (1998).
58. Freeman, K. *et al.* Alterations in cardiac adrenergic signaling and calcium cycling differentially affect the progression of cardiomyopathy. *J. Clin. Invest.* **107**, 967–974 (2001).
59. Akhter, S. A. *et al.* Restoration of β -adrenergic signaling in failing cardiac ventricular myocytes via adenoviral-mediated gene transfer. *Proc. Natl Acad. Sci. USA* **94**, 12100–12105 (1997).
60. Maurice, J. P. *et al.* Enhancement of cardiac function after adenoviral-mediated *in vivo* intracoronary β_2 -adrenergic receptor gene delivery. *J. Clin. Invest.* **104**, 21–29 (1999).
61. Shah, A. S. *et al.* Intracoronary adenovirus-mediated delivery and overexpression of the β_2 -adrenergic receptor in the heart: prospects for molecular ventricular assistance. *Circulation* **101**, 408–414 (2000).
62. Minamisawa, S. *et al.* Chronic phospholamban-sarcoplasmic reticulum calcium ATPase interaction is the critical calcium cycling defect in dilated cardiomyopathy. *Cell* **99**, 313–322 (1999).
63. Roth, D. M. *et al.* Cardiac-directed adenylyl cyclase expression improves heart function in murine cardiomyopathy. *Circulation* **99**, 3099–3102 (1999).
64. Koch, W. J. *et al.* Cardiac function in mice overexpressing the β -adrenergic receptor kinase or a β -ARK inhibitor. *Science* **268**, 1350–1353 (1995).
65. Jaber, M. *et al.* Essential role of β -adrenergic receptor kinase 1 in cardiac development and function. *Proc. Natl Acad. Sci. USA* **93**, 12974–12979 (1996).
66. Rockman, H. A. *et al.* Receptor-specific *in vivo* desensitization by the G protein-coupled receptor kinase-5 in transgenic mice. *Proc. Natl Acad. Sci. USA* **93**, 9954–9959 (1996).
67. Rockman, H. A. *et al.* Control of myocardial contractile function by the level of β -adrenergic receptor kinase 1 in gene-targeted mice. *J. Biol. Chem.* **273**, 18180–18184 (1998).
68. Arber, S. *et al.* MLP-deficient mice exhibit a disruption of cardiac cytoarchitectural organization, dilated cardiomyopathy, and heart failure. *Cell* **88**, 393–403 (1997).
69. Cho, M. C. *et al.* Defective β -adrenergic receptor signaling precedes the development of dilated cardiomyopathy in transgenic mice with calsequestrin overexpression. *J. Biol. Chem.* **274**, 22251–22256 (1999).
70. Freeman, K. *et al.* Progression from hypertrophic to dilated cardiomyopathy in mice that express a mutant myosin transgene. *Am. J. Physiol. Heart Circ. Physiol.* **280**, H151–H159 (2001).
71. Harding, V. B., Jones, L. R., Lefkowitz, R. J., Koch, W. J. & Rockman, H. A. Cardiac β ARK1 inhibition prolongs survival and augments β -blocker therapy in a mouse model of severe heart failure. *Proc. Natl Acad. Sci. USA* **98**, 5809–5814 (2001).
72. White, D. C. *et al.* Preservation of myocardial β -adrenergic receptor signaling delays the development of heart failure after myocardial infarction. *Proc. Natl Acad. Sci. USA* **97**, 5428–5433 (2000).
73. Shah, A. S. *et al.* *In vivo* ventricular gene delivery of a β -adrenergic receptor kinase inhibitor to the failing heart reverses cardiac dysfunction. *Circulation* **103**, 1311–1316 (2001).
74. Naga Prasad, S. V., Barak, L. S., Rapacciuolo, A., Caron, M. G. & Rockman, H. A. Agonist-dependent recruitment of phosphoinositide 3-kinase to the membrane by β -adrenergic receptor Kinase 1. A role in receptor sequestration. *J. Biol. Chem.* **276**, 18953–18959 (2001).
75. The Xamoterol in Severe Heart Failure Study Group. Xamoterol in severe heart failure. *Lancet* **336**, 1–6 (1990).
76. Iaccarino, G., Tomhave, E. D., Lefkowitz, R. J. & Koch, W. J. Reciprocal *in vivo* regulation of myocardial G protein-coupled receptor kinase expression by β -adrenergic receptor stimulation and blockade. *Circulation* **98**, 1783–1789 (1998).
77. McNamara, D. M. *et al.* Pharmacogenetic interactions between β -blocker therapy and the angiotensin-converting enzyme deletion polymorphism in patients with congestive heart failure. *Circulation* **103**, 1644–1648 (2001).

Cardiac channelopathies

Eduardo Marbán

Institute of Molecular Cardiobiology, Ross 844, The Johns Hopkins University, Baltimore, Maryland 21205, USA (e-mail: marban@jhmi.edu)

Genetic alterations of various ion channels produce heritable cardiac arrhythmias that predispose affected individuals to sudden death. The investigation of such 'channelopathies' continues to yield remarkable insights into the molecular basis of cardiac excitability. The concept of channelopathies is not restricted to genetic disorders; notably, changes in the expression or post-translational modification of ion channels underlie the fatal arrhythmias associated with heart failure. Recognizing the fundamental defects in channelopathies provides the basis for new strategies of treatment, including tailored pharmacotherapy and gene therapy.

The heart pumps blood throughout the body and never rests, undergoing roughly three billion cycles in a typical lifetime. To achieve this, the heart must first relax so that its chambers (the atria and ventricles) can fill with blood, and then contract to propel the blood throughout the body. This cycle of relaxation and contraction occurs in a single heartbeat.

Each heartbeat is initiated by a pulse of electrical excitation that begins in a group of specialized pacemaker cells and subsequently spreads throughout the heart. This electrical impulse is made possible by the electrochemical gradient that exists across the surface membrane of each heart cell, or 'myocyte'. At rest, the membrane is selectively permeable to K^+ ions, and the electrochemical potential inside the myocyte is negative with respect to the outside. During electrical excitation, the membrane becomes permeable to Na^+ and the electrochemical potential reverses or 'depolarizes'. Ca^{2+} ions move into the cell and activate the contractile machinery — a process that, when it happens *en masse*, causes the atria and ventricles to contract and expel blood. The wave of depolarization is self-limiting; as a negative membrane potential is restored, the heart relaxes and fills with blood for the next cycle.

Because the heartbeat is so dependent on the proper movement of ions across the surface membrane, disorders of ion channels — or 'channelopathies' — make up a key group of heart diseases. Channelopathies predispose individuals to disturbances of normal cardiac rhythm. If the heart beats too slowly (bradyarrhythmias) or so rapidly that it cannot fill adequately (tachyarrhythmias), then this leads to circulatory collapse and, in the extreme case, death. The incidence of arrhythmias is poorly defined, but conservative estimates are in the range of several million per year in the United States. Arrhythmias lead to more than 250,000 sudden deaths per year, countless lost work days, and financial costs related to treatment including the implantation of more than 250,000 electronic pacemakers and more than 60,000 defibrillators per year. The numbers worldwide are certainly much greater. Several different genetic and acquired channelopathies can cause such arrhythmias.

Genesis of the heartbeat

Normal cardiac excitability results from a delicate balance of depolarizing and repolarizing ionic currents. Each ionic current can be distinguished by its ionic selectivity and time course — properties that are conferred by specific transmembrane proteins called channels or transporters. Figure 1a depicts the principal channels and transporters in heart

cells. Depolarization is mediated by channels or transporters that enable positively charged ions such as Na^+ or Ca^{2+} to enter the cell (or negatively charged ions, usually Cl^- , to exit); the reverse is true for repolarization.

Many proteins have evolved to mediate the selective flux of ions across biological membranes; those that contain a high-throughput pore (more than 100,000 ions per second per molecule) are known as ion channels^{1,2}. Ion channels are generally selective for one type of ion over all others in the physiological milieu. When these channels open, they bias the membrane potential of the cell towards the equilibrium potential of the ion in question. K^+ channels steer the cell towards -90 mV; in contrast, the opening of Na^+ or Ca^{2+} channels forces the potential to positive levels ($+40$ mV or greater). Thus, one can visualize each cell as a dipole that is either positive or negative depending on the relative balance of the cell's complement of ion channels, and whether or not those channels are open at any given time. The resulting electrical signal is known as the 'action potential'.

A prototypical action potential from a ventricular myocyte is shown in the centre of Fig. 1b. The shape is distinctive: a sharp depolarizing upstroke gives way to a sustained, slowly decaying plateau with eventual repolarization. Above the action potential are shown the depolarizing ionic currents, which are generated by the channels (and a single electrogenic transporter, the Na^+/Ca^{2+} exchanger³), that underlie the electrical signal; the nomenclature of the corresponding genes appears on the right. It should be noted that channels allow the passive movement of ions down their respective concentration gradients; when K^+ channels open during repolarization, K^+ exits from the cell. Conversely, during depolarization Na^+ and Ca^{2+} enter the cell. The ionic gradients are maintained ultimately by energy-consuming processes such as the Na^+/K^+ ATPase (see the review in this issue by Bers, pages 198–205).

Na^+ channels and Ca^{2+} channels favour depolarization. Each opens quickly in response to a voltage stimulus (perpetuating further depolarization), and then closes despite maintained depolarization in a process known as 'inactivation'. Under normal conditions, Na^+ channels inactivate quickly and completely, and very rarely re-open⁴. Calcium channels inactivate less rapidly and less completely; they feature prominently in maintaining plateau depolarization⁵.

The molecules involved in the repolarizing mechanism consist of various types of K^+ channels⁶, which are depicted in the lower half of Fig. 1b. The current known as ' I_{K1} ' is active at negative potentials. Its distinctive permeation properties cause I_{K1} to shut-off as depolarization progresses². I_{K1} is therefore ideally suited to anchor the 'resting'

potential, and provides little opposition to depolarization once the balance has tilted in that direction. The other K^+ currents open in distinctive voltage- and time-dependent patterns. The transient outward current ' I_{to} ' has two components: I_{to1} activates and inactivates rapidly, in a conventional voltage-dependent manner^{7,8}; I_{to2} is less well-characterized and varies from species to species, but the available evidence hints that it is activated by changes in the intracellular concentration of free Ca^{2+} (refs 9,10). The *KCND/Kv4* family of channels provides I_{to1} (ref. 11), but the molecular identity of I_{to2} is not yet known. Functionally, I_{to} underlies the 'notch' during the initial phase of repolarization that is often evident in ventricular action potentials¹², as shown in Fig. 1b. I_{to} also influences the overall duration of the action potential, albeit indirectly^{13,14}.

The delayed rectifier I_K , like I_{to} , consists of two components: a rapid component (I_{Kr}) and a slow component (I_{Ks}). Mutations in each of the four genes encoding I_{Kr} and I_{Ks} have been implicated in heritable long-QT syndrome¹⁵, as discussed below. Finally, I_{Kp} is a time-independent background current¹⁶ whose molecular identity

is uncertain but may be related to two-pore K^+ channels of the *KCNK* family^{17,18}. Computational simulations have shown that I_{Kp} is essential for repolarizing action potentials¹⁹, but it remains to be studied extensively at a biological level.

The electrocardiogram

The mass of the heart is sufficient to allow electrodes placed on the body surface to register readily the electrical changes generated by cardiac myocytes. The resultant signal — the electrocardiogram — represents an average of the electrical gradients at any given time. Figure 2a shows the temporal relationships between the electrical activity of a typical ventricular myocyte, as would be measured using cellular recordings of the transmembrane potential, and the corresponding electrocardiogram.

The upstroke of the action potential at the onset of depolarization produces the spiky 'QRS complex'; repolarization is manifested on the body surface as the gently rolling T wave. As a first approximation, the time between the beginning of the QRS complex and the end of the T wave — the 'QT interval' — can be used to deduce the overall timing and duration of ventricular depolarization and repolarization. The frequency of QRS complexes and their sequence relative to the smaller P waves produced by atrial activity allow the clinical detection of normal rhythm (Fig. 2b) or rhythm disorders (see below).

Mechanism of arrhythmias

Repolarization of the heart cell is a precarious process²⁰. Although many different currents orchestrate repolarization, only a few channels of each type are open at any given time; thus, small changes, corresponding to the opening or closing of a handful of individual channel molecules²¹, can pervert repolarization. The normally smooth trajectory of repolarization might then be interrupted by abnormal, secondary depolarizations²². Such 'afterdepolarizations'²³ are innocuous in isolated cells, but in the syncytial heart, myocytes are coupled to their neighbours, enabling the electrical impulse to spread from cell to cell (ref. 24; and Fig. 3a).

Imagine that one region of the heart undergoes afterdepolarizations while neighbouring regions have begun to repolarize (Fig. 3b). The electrocardiogram would show a prolongation of the QT interval. The aberrant zone may then re-excite its repolarizing neighbours, initiating a premature beat that spreads in wavelets throughout the heart^{25,26}. Self-perpetuating wavelets of excitation in the ventricle undermine the normal, stereotyped progression of the electrical impulse, producing ventricular tachycardia (Fig. 2c, large wavelets) or ventricular fibrillation (small wavelets)²⁷. The resultant incoordinate contraction and rapid rate lead quickly to circulatory collapse and death if the arrhythmia is sustained.

Heritable channelopathies

Disorders of repolarization

Abnormal, inhomogeneous repolarization underlies the ventricular arrhythmias characteristic of many cardiac channelopathies. The inherent instability of cardiac repolarization is accentuated when action potentials become prolonged²³, leading to a characteristic prolongation of the QT interval. Thus, the 'long-QT syndrome' — the archetypal disease of repolarization — takes its name from the distinctive abnormality of the electrocardiogram in affected individuals. Such individuals are predisposed to ventricular tachyarrhythmias caused by unstable repolarization, which often leads to sudden cardiac death as the first manifestation of the disease.

Long-QT syndrome can be heritable or acquired²⁸. The genetically transmitted form is caused by discrete mutations in genes that encode ion channels. Either the Na^+ channel gene *SCN5A* or one of the four genes that make up I_K can be affected¹⁵. (The existence of other long-QT genes is suggested by linkage analysis, but their identity is not yet known.) Long-QT-associated mutations in *SCN5A* produce channels with increased Na^+ flux²⁹, whereas those in the K^+ channels lead to loss of function³⁰. The common mechanistic thread is perturbation of the

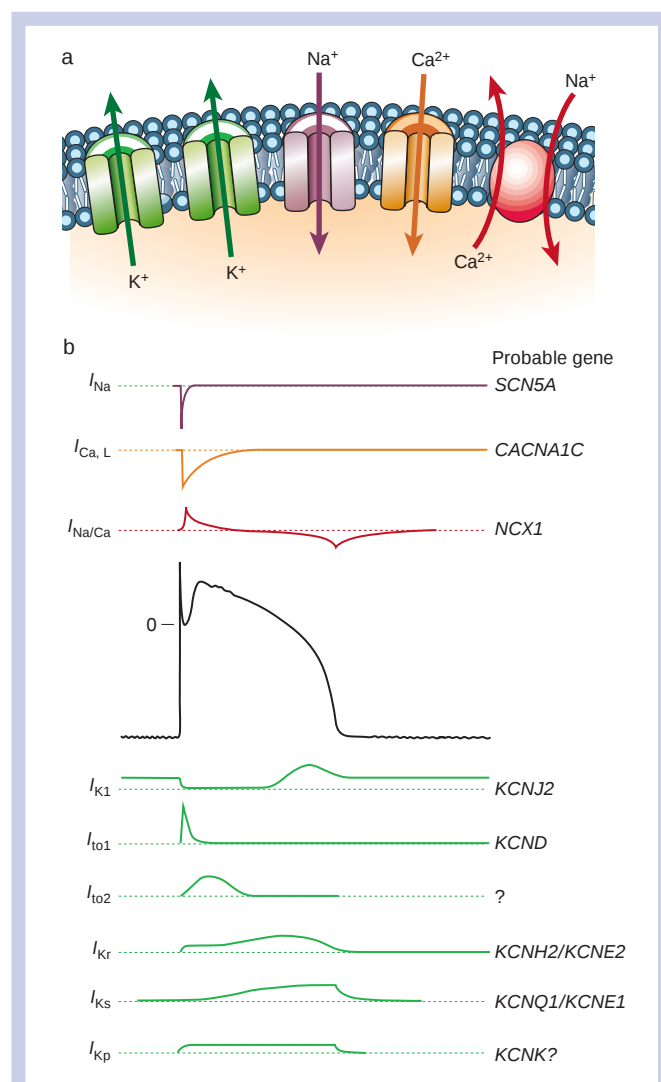


Figure 1 Ion channels underlie cardiac excitability. **a**, The key ion channels (and an electrogenic transporter) in cardiac cells. K^+ channels (green) mediate K^+ efflux from the cell; Na^+ channels (purple) and Ca^{2+} channels (yellow) mediate Na^+ and Ca^{2+} influx, respectively. The Na^+/Ca^{2+} exchanger (red) is electrogenic, as it transports three Na^+ ions for each Ca^{2+} ion across the surface membrane. **b**, Ionic currents and genes underlying the cardiac action potential. Top, depolarizing currents as functions of time, and their corresponding genes; centre, a ventricular action potential; bottom, repolarizing currents and their corresponding genes.

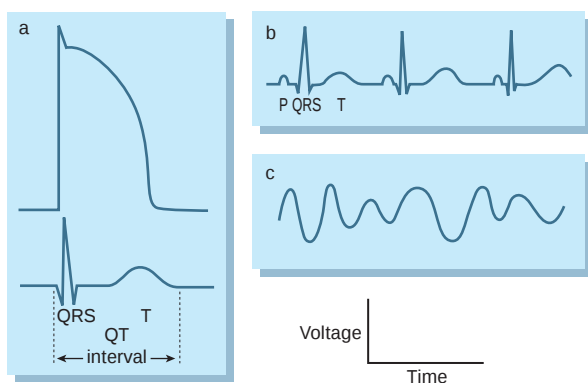


Figure 2 The action potential and the electrocardiogram (ECG). **a**, Temporal relationships between the ventricular action potential (top) and the ECG (bottom). The QRS complex, T wave and QT interval are indicated. Both signals are functions of the same timescale on the *x* axis; the *y* axis plots voltage, with a gain that is roughly 100-fold higher for the ECG than for the action potential (that is, the absolute signal amplitude is about 100-fold smaller for the ECG). **b**, Normal rhythm on the ECG. P waves produced by atrial activity precede each QRS complex. **c**, Ventricular tachycardia. QRS complexes are broadened and irregular; independent or retrograde P wave activity may be evident (not shown here).

balance between inward and outward currents during the plateau of the action potential. The inheritance pattern is most frequently autosomal dominant: alterations of a single allele are sufficient to produce the arrhythmogenic phenotype. Functional studies of defined channel mutations in heterologous expression systems or *in vivo* have helped to rationalize the pathophysiology. *SCN5A* mutations typically yield incompletely inactivating Na^+ channels³¹, leading to a tendency towards unbalanced depolarizing throughout the action potential plateau. This gain of function explains why one abnormal allele is sufficient to undermine repolarization.

In contrast, long-QT-associated mutations in the K^+ channels decrease K^+ flux through I_{Kr} or I_{Ks} by loss-of-function (for example, nonsense mutations that truncate the pore-forming subunit prematurely) or dominant-negative mechanisms^{30,32}. Because K^+ channels are multimeric (unlike Na^+ channels, in which a single protein is sufficient to create a functional pore), the dominant-negative mutations cripple the healthy products of the wild-type allele, and thus provide a ready rationale for dominant transmission. The fact that plain loss-of-function mutations also produce dominantly inherited long-QT syndrome implies, however, that two functional alleles are required for uneventful repolarization. Assuming that the loss of only one functional allele will produce no more than a 50% reduction of either I_{Kr} or I_{Ks} , these findings illustrate vividly the precarious nature of the repolarization process.

More rarely, long-QT syndrome is inherited in an autosomal recessive manner. Such kindreds possess two dysfunctional K^+ channel genes, which leads to a total absence of I_{Kr} or I_{Ks} (ref. 33). Individuals also affected with associated deafness (Jervell and Lange-Nielsen syndrome) have mutations in the I_{Ks} genes *KCNQ1* or *KCNE1*; indeed, studies motivated by such individuals led to the recognition that I_{Ks} is necessary for endolymph production in the inner ear¹⁵. Much has been made of other, more subtle genotype-phenotype correlations in long-QT syndrome, such as gene-specific variations in the precise electrocardiographic features³⁴; however, the paucity of people with these rare disorders weakens such correlations. It should also be noted that there is extensive phenotypic variability among known gene carriers, even within the same family, hinting that there are modifier genes yet to be recognized.

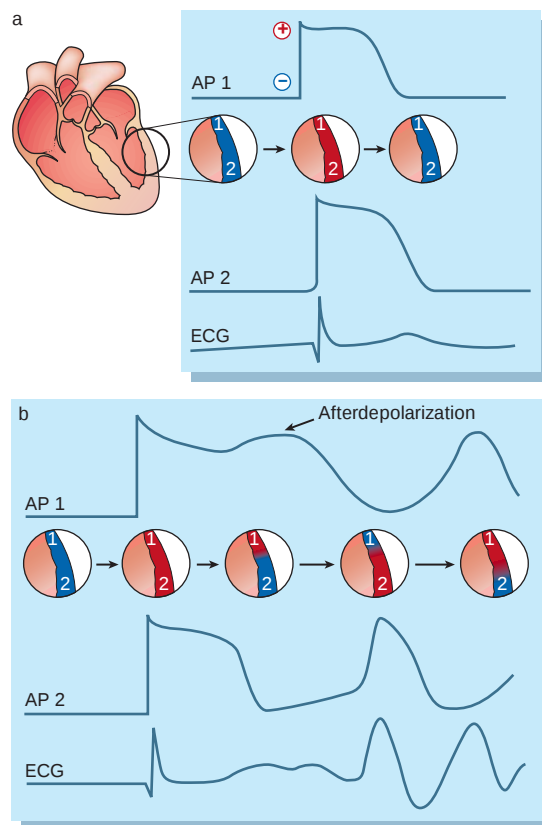


Figure 3 Explanation of the mechanism underlying arrhythmia in long-QT syndrome. **a**, Time series of a cross-section of the ventricular wall in normal myocardium. Negatively polarized (resting) muscle is shown as blue; depolarized muscle is red. Action potentials recorded at sites 1 and 2 are similar in timing and morphology. The resulting ECG is normal (compare Fig. 2a). **b**, Time series of a cross-section of the ventricular wall in long-QT myocardium. Part of the ventricular wall undergoes an afterdepolarization (site 1), whereas neighbouring site 2 has repolarized. The afterdepolarization at site 1 prematurely re-excites site 2, initiating ventricular tachycardia.

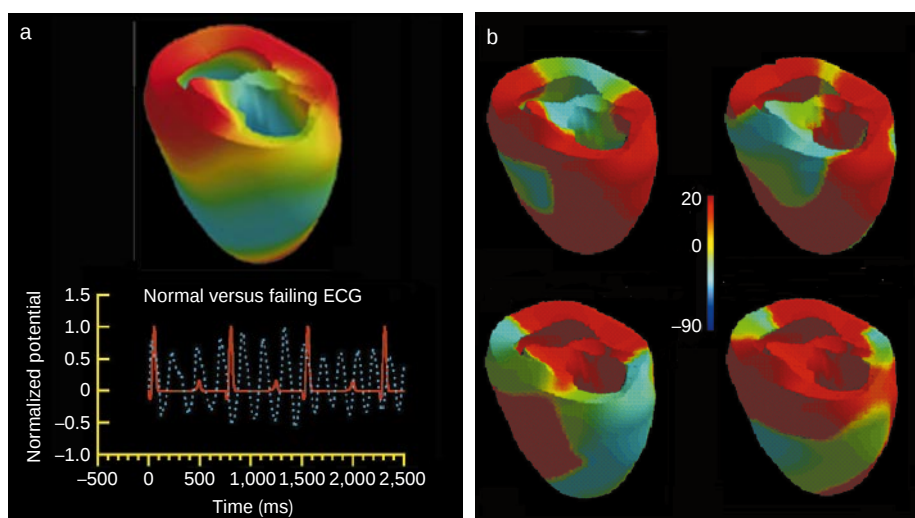
With the exception of Jervell and Lange-Nielsen syndrome, with its associated deafness, the classical long-QT syndrome is notable for its restriction to the heart. Although the affected K^+ channel genes are expressed in various tissues, the phenotype hints that the physiological roles of these genes are most crucial in the heart. In contrast, a rare genetic disease known as Andersen's syndrome, in which long-QT syndrome is associated with multi-system pathology (periodic paralysis and dysmorphisms), has been attributed to mutations in *KCNJ2* (ref. 35). Mutations in various kindreds and sporadic cases occur throughout the *KCNJ2* coding sequence; two mutants that have been characterized functionally have dominant-negative properties, consistent with the observed autosomal dominant inheritance pattern. These findings indicate that *KCNJ2* is functionally important not only in stabilizing cardiac rhythm, but also in modulating the excitability of skeletal muscle and in morphogenesis².

Loss of Na^+ channel function and arrhythmia

Another uncommon but instructive cardiac channelopathy is that of idiopathic ventricular fibrillation, in which previously well individuals die suddenly of a tachyarrhythmia³⁶. The electrocardiogram can be normal at baseline, although some individuals (with so-called Brugada syndrome) have associated electrocardiographic abnormalities (including a form of intraventricular conduction delay called right bundle branch block)³⁷. The disease has been linked convincingly to

Figure 4 Computational rationalization of heart failure arrhythmias on the basis of known cellular changes in ionic currents.

a, Simulation of normal rhythm in the heart, with the wave of excitation (red) spreading rapidly throughout the ventricles. The resulting virtual electrocardiogram closely resembles normal rhythm (bottom trace, solid red line). **b**, Four snapshots during ventricular tachyarrhythmia in a virtual heart that contains a region of cells with properties of heart failure (a form of acquired long-QT syndrome). The calculated electrocardiogram reproduces ventricular tachycardia (superimposed on the normal rhythm shown in **a** as a dashed line). Results shown are from R. Winslow, Johns Hopkins University.



SCN5A with dominant inheritance^{36,38}. Interestingly, the functional abnormalities of the expressed mutant channels are opposite to those found in Na⁺ channel mutants associated with long-QT syndrome. In idiopathic ventricular fibrillation, the Na⁺ channels show loss-of-function features such as enhanced inactivation. Affected channels may not even make it to the surface membrane at physiological temperatures³⁹; thus, the available evidence suggests that the syndrome is one of pure or partial hemi-allelic Na⁺ channel insufficiency.

There is no intuitive rationale for ventricular arrhythmias caused by partial Na⁺ channel deficiency. The principal physiological role of Na⁺ channels is the fast conduction of the cardiac impulse⁴⁰. What, then, is predisposing the heart to fatal tachyarrhythmias? One hypothesis is that the key abnormality lies, once again, in the repolarization process. The premise is as follows: Na⁺ channels inactivate quickly, but before doing so they initiate the action potential plateau, and sustain it for several milliseconds until Ca²⁺ channels have a chance to function. I_{to} , which is richly expressed in the outer layers (epicardium) of the human ventricle¹², simultaneously turns on and opposes the depolarization (normally producing the action potential notch). Thus, the early plateau represents a tug of war between Na⁺ channels and I_{to} . A decreased density of Na⁺ channels in the face of a robust I_{to} may allow premature repolarization, resulting in a very brief action potential. If this occurs more readily in the epicardium, where I_{to} is prominent, than in the inner ventricular layers, then the normally depolarized inner layers can re-excite the prematurely repolarized epicardium. The essential feature is the postulated inhomogeneity of repolarization across the thickness of the ventricular wall³⁷, which is brought about by an imbalance between I_{Na} and I_{to} .

Although plausible, transmural inhomogeneity may not explain fully the ventricular arrhythmias associated with Na⁺ channel deficiency. Some individuals affected with complete hemi-allelic Na⁺ channel insufficiency present only with isolated slowing of myocardial conduction⁴¹; other individuals with less severe Na⁺ channel gating defects than those typically seen in Brugada syndrome likewise show isolated conduction slowing⁴². Thus, although we recognize that the full spectrum of syndromes linked to Na⁺ channel deficiency includes conduction slowing and ventricular arrhythmias⁴³, much remains to be clarified about the detailed genotype–phenotype correlations.

Acquired channelopathies

Heart failure

The heritable channelopathies have yielded important insights into the pathophysiology of some far more common, acquired diseases. Heart failure is a case in point. This disease afflicts hundreds of

millions of people worldwide. Whatever the initiating factors (for example, coronary atherosclerosis, hypertension or viral infection), the final common phenotype is one of a dilated, poorly contracting heart. Affected individuals suffer from a decreased ability to exercise and shortness of breath caused by a decrease in cardiac pump function. Mortality remains high despite the best current therapy, exceeding 10% per year in severely symptomatic individuals. Although the name 'heart failure' suggests that gradually dwindling cardiac output might be the most likely cause of death, it turns out that most individuals die suddenly of cardiac arrhythmias.

We now know that heart failure represents a common, acquired form of the long-QT syndrome⁴⁴. Myocytes from failing hearts show prolongation of action potentials^{45,46}, and repolarization *in vivo* is abnormally labile⁴⁷. In human heart failure, the action potential prolongation reflects selective downregulation of two K⁺ currents, I_{to1} and I_{K1} (ref. 45). Much of the decrease in I_{to1} occurs at the transcriptional level¹¹. Such K⁺ channel downregulation may be adaptive in the short term: increased depolarization during the cardiac cycle means more time is available for excitation–contraction coupling, which mitigates the decrease in cardiac output. Nevertheless, the downregulation of K⁺ channels becomes maladaptive in the long term, predisposing the individual to afterdepolarizations, inhomogeneous repolarization and ventricular tachyarrhythmias. The arrhythmic tendency is aggravated by alterations in the cycling of intracellular calcium, including upregulation of the electrogenic Na⁺/Ca²⁺ exchanger (ref. 48, and see review in this issue by Bers, pages 198–205).

Factually based numerical models of electrical activity have begun to shed light on the mechanisms of cardiac arrhythmias. The initial successes came in simulations carried out on a cellular level^{19,49}, which rationalized the mechanisms of long-QT-related action potential prolongation and afterdepolarizations⁵. More recently, massively parallel network simulations of whole-heart electrical activity have reproduced successfully polymorphic ventricular tachycardia — an arrhythmia commonly seen in heart failure. The model¹⁴ includes virtual cells coupled to each other in a geometry defined from an actual mammalian heart; each cell contains a biophysically detailed model of cardiac electrophysiology, including most of the ionic currents listed in Fig. 1.

The results of the whole-heart simulation are highlighted in Fig. 4 and in video clips available at <http://www.cmbl.jhu.edu/movies/normal.mpg> and <http://www.cmbl.jhu.edu/movies/ead.mpg>. Normal conduction and repolarization can be reproduced in normal hearts (Fig. 4a); in contrast, when an area of the ventricle is reprogrammed

to reproduce the selective downregulation of K^+ channels characteristic of heart failure, an arrhythmia can be induced readily *in silico* (Fig. 4b). The initiating event is an afterdepolarization from the K^+ -channel-deficient zone; once triggered, the arrhythmia is perpetuated by rapid waves of depolarization spreading asynchronously throughout the heart. Virtual electrocardiograms reveal the undulating waveform of polymorphic ventricular tachycardia (compare the simulated electrograms in Fig. 4a with those in Fig. 2b,c). The biological hypothesis of repolarization-related arrhythmias has thus been validated numerically from first principles. Given the complexity of cardiac arrhythmias, such *in silico* simulations will undoubtedly feature more prominently in future investigations.

Drug-induced long-QT syndrome

In addition to heart failure, acquired long-QT syndrome can also be induced by exposure to drugs that block K^+ channels^{28,50}. Selective blockers of I_{Kr} , such as dofetilide have been developed for the treatment of various atrial arrhythmias; unfortunately, such drugs predictably evoke prolongation of the QT interval, which is sufficient to cause dangerous ventricular arrhythmias in 5–7% of recipients. Many other drugs block K^+ channels unintentionally. I_{Kr} is a common target — a fact that can be interpreted through the unique structural features of the *KCNH2* inner vestibule that render it rather promiscuous for small organic molecules⁵¹. Drug-induced arrhythmias occur more frequently in women than in men, and in a small percentage of those exposed to the drugs in question⁵². In these individuals, however, the arrhythmia can be lethal; such side-effects have led to the withdrawal of various prescription medications after their initial approval by regulatory agencies.

It is not yet known why a subpopulation is particularly prone to drug-induced long-QT syndrome. One notion is that the repolarization process has a certain 'reserve' built in by virtue of the redundancy of K^+ channels and their normal levels of expression. A diminished repolarization reserve, perhaps caused by a channel mutation that alone does not cause symptoms, may potentially lead to arrhythmias in the presence of certain drugs⁵³. In particular, otherwise innocent polymorphisms in ion channel genes may enhance drug binding and magnify the channel block⁵⁴. The prototype for such polymorphisms may be a mutation in *KCNE2* that has been reported to underlie arrhythmias triggered by an antibiotic⁵⁵, but the findings are so far restricted to a single proband. Indeed, the mechanism whereby *KCNE2* mutations produce repolarization instability remains uncertain⁵⁶. The basis of gender differences is also unknown, but is consistent with the fact that the QT interval is longer, at baseline, in women than in men⁵².

The future in diagnosis and treatment

The clinical criteria for diagnosing long-QT syndrome and other cardiac channelopathies remain the traditional ones: history, physical examination and electrocardiography⁵⁷. The alternatives for genetic diagnosis are linkage analysis, which is practicable only in large families, and a candidate gene approach. No commercial high-throughput approach to genetic diagnosis is available as yet; indeed, such tests may be some time in coming, given the rarity of the classically inherited syndromes, the ever-increasing number of documented mutations, and the likelihood that several other culpable genes have yet to be recognized. Nevertheless, those individuals who have been genotyped may benefit from tailored pharmacotherapy.

Long-QT syndrome caused by mutations in *SCN5A* can be treated rationally using Na^+ channel blockers of the local anaesthetic class, such as mexilitine. Such drugs preferentially block the non-inactivating mutant Na^+ channel current; their clinical value has been documented in various case reports^{58,59}. The existing pharmacopeia is less helpful in long-QT syndrome related to K^+ channels. Therapeutic measures centre on general measures such as oral potassium supplementation⁶⁰, strict avoidance of aggravating factors such as

potassium-wasting diuretics, and the implantation of electronic devices to regularize rhythm (pacemakers and/or automatic cardioverter/defibrillators). Other avenues that merit further investigation include K^+ channel agonist drugs⁶¹ and gene therapy.

Gene therapy for cardiac arrhythmias?

The currently available therapies for arrhythmias are limited by poor efficacy and the incidence of serious side-effects. Options for treatment include pharmacotherapy, radiofrequency ablation and implanted devices. Antiarrhythmic medications can sometimes reduce primary arrhythmic events, but their systemic effects are often poorly tolerated; in addition, their paradoxical ability to make some arrhythmias worse while treating others actually increases mortality in many situations. Radiofrequency ablation cures a limited number of arrhythmias and has become the standard of care for people with congenital structural abnormalities (for example, Wolff–Parkinson–White syndrome), but more common atrial and ventricular tachycardias are less amenable to this form of therapy. Device-based therapies (pacemakers and defibrillators), while palliative, can be quite effective. However, this strategy does not prevent tachyarrhythmias and is associated with a lifetime commitment to repeated procedures, significant expense and potentially catastrophic complications (lung or heart perforation, lead dislodgement, or infection).

The lack of effective therapeutic options motivates the pursuit of alternative strategies for cardiac arrhythmias — notably gene therapy. The philosophy is very different from that of conventional gene therapy (see review in this issue by Isner, pages 234–239); here, the purpose of gene transfer is to effect electrical re-engineering. The genes in question are those that encode either ion channels or modulators of ion channels such as G proteins (see review in this issue by Rockman, Koch and Lefkowitz, pages 206–212). The concepts can be generalized to ventricular arrhythmias, such as those discussed above. In heart failure, for example, an overexpression of K^+ channels can be used to antagonize acquired long-QT syndrome^{62,63}, and the attendant loss of contractility may be amenable to co-administration of a second gene to augment calcium cycling, in a strategy of dual gene therapy. Although such work is conceptually attractive, widespread delivery with long-term expression will be required before human trials can be anticipated.

More appealing targets for development in the short term are arrhythmias in which very local modifications of electrical properties are sufficient for effective treatment. An example of such an arrhythmia is atrial fibrillation, which results from the rapid and uncoordinated firing of electrical impulses from several sites in the atria (see review in this issue by Nattel, pages 219–226). Many of these impulses then travel to the ventricles, resulting in irregular, erratic and rapid heart rhythm. Atrial fibrillation affects more than two million people in the United States alone. Initially, treatment is directed at maintaining normal sinus rhythm, but this is rarely successful in the long term. When atrial fibrillation becomes persistent, therapy focuses on controlling the ventricular rate. Such rate control can be achieved by local modification of electrical conduction in the atrioventricular node, because the atrioventricular node is the only pathway for conducting the electrical impulse from the rapidly activating atria to the ventricles. Donahue *et al.*⁶⁴ have shown the feasibility of atrioventricular nodal modification by gene therapy in an animal model of atrial fibrillation. By locally overexpressing an inhibitory G protein (G_i) subunit, these investigators could slow atrioventricular conduction without affecting other electrical parameters. The therapy resulted in a salutary reduction of the ventricular rate during atrial fibrillation. Local gene delivery was highly enriched by selective infusion of the transgenes into the branch of the coronary arteries that supplies the atrioventricular node, and accomplished percutaneously using clinically available catheters.

There are singular advantages of gene therapy for atrial fibrillation. First, highly localized gene delivery is sufficient to treat the

problem. The amount of gene delivered can be reduced correspondingly, and potential problems owing to widespread dissemination can be averted more readily. Second, treated cells remain responsive to endogenous nerves and hormones. Such was the case with overexpression of G_i in the atrioventricular node: atrioventricular conduction remained responsive to β -adrenergic stimulation. Third, implanted hardware is avoided, obviating long-term risks and the expense and morbidity associated with battery and lead replacements. Fourth, the localized coronary circulation allows isolated delivery to the atrioventricular node. Fifth, the proximity to the inner lining of the heart, the endocardium, allows access by intracardiac injection, providing a potential alternative delivery route. Sixth, the therapeutic effects can be readily detected by electrocardiography. Last, changes induced by gene transfer can be rescued by conventional electrophysiological methods (atrioventricular node ablation and pacemaker implantation).

Similar advantages of gene therapy may be exploitable in the creation of genetically engineered pacemakers; such a possibility might provide the first viable alternative to electronic pacemakers for the treatment of bradyarrhythmias. Thus, many arrhythmias may turn out to be reasonable targets for functional re-engineering by gene transfer. This concept, and several others reviewed above — *in silico* modelling approaches, tailored drug therapy and pharmacogenomics — give ample reason to hope that channelopathies will not only be better understood in the future, but will also become increasingly amenable to rational therapy. □

- Catterall, W. A. Molecular properties of sodium and calcium channels. *J. Bioenerg. Biomembr.* **28**, 219–230 (1996).
- Jan, L. Y. & Jan, Y. N. Voltage-gated and inwardly rectifying potassium channels. *J. Physiol.* **505**, 267–282 (1997).
- Philipson, K. D. & Nicoll, D. A. Sodium–calcium exchange: a molecular perspective. *Annu. Rev. Physiol.* **62**, 111–133 (2000).
- Marban, E., Yamagishi, T. & Tomaselli, G. F. Structure and function of voltage-gated sodium channels. *J. Physiol.* **508**, 647–657 (1998).
- Zeng, J. & Rudy, Y. Early afterdepolarizations in cardiac myocytes: mechanism and rate dependence. *Biophys. J.* **68**, 949–964 (1995).
- Robbins, J. KCNQ channels: physiology, pathophysiology, and pharmacology. *Pharmacol. Ther.* **90**, 1–19 (2001).
- Näbauer, M., Beuckelmann, D. J. & Erdmann, E. Characteristics of transient outward current in human ventricular myocytes from patients with terminal heart failure. *Circ. Res.* **73**, 386–394 (1993).
- Hoppe, U. C. *et al.* Manipulation of cellular excitability by cell fusion: effects of rapid introduction of transient outward K^+ current on the guinea pig action potential. *Circ. Res.* **84**, 964–972 (1999).
- Papp, Z. *et al.* Two components of $[Ca^{2+}]_i$ -activated Cl^- current during large $[Ca^{2+}]_i$ transients in single rabbit heart Purkinje cells. *J. Physiol.* **483**, 319–330 (1995).
- Koster, O. F., Sziget, G. P. & Beuckelmann, D. J. Characterization of a $[Ca^{2+}]_i$ -dependent current in human atrial and ventricular cardiomyocytes in the absence of Na^+ and K^+ . *Cardiovasc. Res.* **41**, 175–187 (1999).
- Käbb, S. *et al.* Molecular basis of transient outward potassium current downregulation in human heart failure: a decrease in Kv4.3 mRNA correlates with a reduction in current density. *Circulation* **98**, 1383–1393 (1998).
- Näbauer, M. *et al.* Regional differences in current density and rate-dependent properties of the transient outward current in subepicardial and subendocardial myocytes of human left ventricle. *Circulation* **93**, 168–177 (1996).
- Hoppe, U. C., Marban, E. & Johns, D. C. Molecular dissection of cardiac repolarization by *in vivo* Kv4.3 gene transfer. *J. Clin. Invest.* **105**, 1077–1084 (2000).
- Winslow, R. L., Rice, J. & Jafri, S. Modeling the cellular basis of altered excitation–contraction coupling in heart failure. *Prog. Biophys. Mol. Biol.* **69**, 497–514 (1998).
- Keating, M. T. & Sanguinetti, M. C. Molecular and cellular mechanisms of cardiac arrhythmias. *Cell* **104**, 569–580 (2001).
- Yue, D. T. & Marban, E. A novel cardiac potassium channel that is active and conductive at depolarized potentials. *Pflügers Arch.* **413**, 127–133 (1988).
- Patel, A. J., Lazdunski, M. & Honore, E. Lipid and mechano-gated 2P domain K^+ channels. *Curr. Opin. Cell Biol.* **13**, 422–428 (2001).
- Kim, D. *et al.* Cloning and functional expression of a novel cardiac two-pore background K^+ channel (cTBK-1). *Circ. Res.* **82**, 513–518 (1998).
- Luo, C. H. & Rudy, Y. A dynamic model of the cardiac ventricular action potential. I. Simulations of ionic currents and concentration changes. *Circ. Res.* **74**, 1071–1096 (1994).
- Weidmann, S. The electrical constants of Purkinje fibres. *J. Physiol.* **118**, 348–360 (1952).
- Nichols, C. G., Ripoll, C. & Lederer, W. J. ATP-sensitive potassium channel modulation of the guinea pig ventricular action potential and contraction. *Circ. Res.* **68**, 280–287 (1991).
- Marban, E., Robinson, S. W. & Wier, W. G. Mechanisms of arrhythmogenic delayed and early afterdepolarizations in ferret ventricular muscle. *J. Clin. Invest.* **78**, 1185–1192 (1986).
- Cranefield, P. F. & Aronson, R. S. Torsades de pointes and early afterdepolarizations. *Cardiovasc. Drugs Ther.* **5**, 531–537 (1991).
- Saffitz, J. E., Laing, J. G. & Yamada, K. A. Connexin expression and turnover: implications for cardiac excitability. *Circ. Res.* **86**, 723–728 (2000).
- Berenfeld, O. & Jalife, J. Purkinje–muscle reentry as a mechanism of polymorphic ventricular arrhythmias in a 3-dimensional model of the ventricles. *Circ. Res.* **82**, 1063–1077 (1998).
- Hoffman, B. F. & Rosen, M. R. Cellular mechanisms for cardiac arrhythmias. *Circ. Res.* **49**, 1–15 (1981).
- Samie, F. H. & Jalife, J. Mechanisms underlying ventricular tachycardia and its transition to ventricular fibrillation in the structurally normal heart. *Cardiovasc. Res.* **50**, 242–250 (2001).
- Camm, A. J. *et al.* Congenital and acquired long QT syndrome. *Eur. Heart J.* **21**, 1232–1237 (2000).
- Wang, Q. *et al.* SCN5A mutations associated with an inherited cardiac arrhythmia, long QT syndrome. *Cell* **80**, 805–811 (1995).
- Sanguinetti, M. C. Dysfunction of delayed rectifier potassium channels in an inherited cardiac arrhythmia. *Ann. NY Acad. Sci.* **868**, 406–413 (1999).
- Dumaine, R. *et al.* Multiple mechanisms of Na^+ channel-linked long-QT syndrome. *Circ. Res.* **78**, 916–924 (1996).
- Hoppe, U. C., Marban, E. & Johns, D. C. Distinct gene-specific mechanisms of arrhythmia revealed by cardiac gene transfer of two long QT disease genes, HERG and KCNE1. *Proc. Natl Acad. Sci. USA* **98**, 5335–5340 (2001).
- Spawski, I. *et al.* Spectrum of mutations in long-QT syndrome genes. KVLQT1, HERG, SCN5A, KCNE1, and KCNE2. *Circulation* **102**, 1178–1185 (2000).
- Schwartz, P. J. *et al.* Genotype–phenotype correlation in the long-QT syndrome: gene-specific triggers for life-threatening arrhythmias. *Circulation* **103**, 89–95 (2001).
- Plaster, N. M. *et al.* Mutations in Kir2.1 cause the developmental and episodic electrical phenotypes of Andersen's syndrome. *Cell* **105**, 511–519 (2001).
- Chen, Q. *et al.* Genetic basis and molecular mechanism for idiopathic ventricular fibrillation. *Nature* **392**, 293–296 (1998).
- Antzelevitch, C. The Brugada syndrome: ionic basis and arrhythmia mechanisms. *J. Cardiovasc. Electrophysiol.* **12**, 268–272 (2001).
- Rook, M. B. *et al.* Human SCN5A gene mutations alter cardiac sodium channel kinetics and are associated with the Brugada syndrome. *Cardiovasc. Res.* **44**, 507–517 (1999).
- Dumaine, R. *et al.* Ionic mechanisms responsible for the electrocardiographic phenotype of the Brugada syndrome are temperature dependent. *Circ. Res.* **85**, 803–809 (1999).
- Wang, Y. & Rudy, Y. Action potential propagation in inhomogeneous cardiac tissue: safety factor considerations and ionic mechanism. *Am. J. Physiol. Heart Circ. Physiol.* **278**, H1019–H1029 (2000).
- Schott, J. J. *et al.* Cardiac conduction defects associate with mutations in SCN5A. *Nature Genet.* **23**, 20–21 (1999).
- Tan, H. L. *et al.* A sodium-channel mutation causes isolated cardiac conduction disease. *Nature* **409**, 1043–1047 (2001).
- Veldkamp, M. W. *et al.* Two distinct congenital arrhythmias evoked by a multidysfunctional Na^+ channel. *Circ. Res.* **86**, E91–E97 (2000).
- Marban, E. Heart failure: the electrophysiologic connection. *J. Cardiovasc. Electrophysiol.* **10**, 1425–1428 (1999).
- Beuckelmann, D. J., Näbauer, M. & Erdmann, E. Alterations of K^+ currents in isolated human ventricular myocytes from patients with terminal heart failure. *Circ. Res.* **73**, 379–385 (1993).
- Käbb, S. *et al.* Ionic mechanism of action potential prolongation in ventricular myocytes from dogs with pacing-induced heart failure. *Circ. Res.* **78**, 262–273 (1996).
- Berger, R. D. *et al.* Beat-to-beat QT interval variability: novel evidence for repolarization lability in ischemic and nonischemic dilated cardiomyopathy. *Circulation* **96**, 1557–1565 (1997).
- Studer, R. *et al.* Gene expression of the cardiac Na^+Ca^{2+} exchanger in end-stage human heart failure. *Circ. Res.* **75**, 443–453 (1994).
- Winslow, R. L. *et al.* Mechanisms of altered excitation–contraction coupling in canine tachycardia-induced heart failure. II: model studies. *Circ. Res.* **84**, 571–586 (1999).
- Sanguinetti, M. C. *et al.* A mechanistic link between an inherited and an acquired cardiac arrhythmia: HERG encodes the I_{Kr} potassium channel. *Cell* **81**, 299–307 (1995).
- Mitcheson, J. S. *et al.* A structural basis for drug-induced long QT syndrome. *Proc. Natl Acad. Sci. USA* **97**, 12329–12333 (2000).
- Liu, X. K. *et al.* Female gender is a risk factor for torsades de pointes in an *in vitro* animal model. *J. Cardiovasc. Pharmacol.* **34**, 287–294 (1999).
- Roden, D. M. & Spooner, P. M. Inherited long QT syndromes: a paradigm for understanding arrhythmogenesis. *J. Cardiovasc. Electrophysiol.* **10**, 1664–1683 (1999).
- Roden, D. M. Pharmacogenetics and drug-induced arrhythmias. *Cardiovasc. Res.* **50**, 224–231 (2001).
- Abbott, G. W. *et al.* MiRP1 forms I_{Kr} potassium channels with HERG and is associated with cardiac arrhythmia. *Cell* **97**, 175–187 (1999).
- Mazhari, R. *et al.* Molecular interactions between two long-QT syndrome gene products, HERG and KCNE2, rationalized by *in vitro* and *in silico* analysis. *Circ. Res.* **89**, 33–38 (2001).
- Priori, S. G. *et al.* Genetic and molecular basis of cardiac arrhythmias: impact on clinical management parts I and II. *Circulation* **99**, 518–528 (1999).
- Benhorin, J. *et al.* Effects of flecainide in patients with new SCN5A mutation: mutation-specific therapy for long-QT syndrome? *Circulation* **101**, 1698–1706 (2000).
- Windle, J. R. *et al.* Normalization of ventricular repolarization with flecainide in long QT syndrome patients with SCN5A:DeltaKPQ mutation. *Ann. Noninvasive Electrocardiol.* **6**, 153–158 (2001).
- Compton, S. J. *et al.* Genetically defined therapy of inherited long-QT syndrome. Correction of abnormal repolarization by potassium. *Circulation* **94**, 1018–1022 (1996).
- Shimizu, W. *et al.* Improvement of repolarization abnormalities by a K^+ channel opener in the LQT1 form of congenital long-QT syndrome. *Circulation* **97**, 1581–1588 (1998).
- Nuss, H. B. *et al.* Reversal of potassium channel deficiency in cells from failing hearts by adenoviral gene transfer: a prototype for gene therapy for disorders of cardiac excitability and contractility. *Gene Ther.* **3**, 900–912 (1996).
- Nuss, H. B., Marban, E. & Johns, D. C. Overexpression of a human potassium channel suppresses cardiac hyperexcitability in rabbit ventricular myocytes. *J. Clin. Invest.* **103**, 889–896 (1999).
- Donahue, J. K. *et al.* Focal modification of electrical conduction in the heart by viral gene transfer. *Nature Med.* **6**, 1395–1398 (2000).

Acknowledgements

Supported by NIH. E.M. is the Michel Mirowski, M.D. Professor of Cardiology of the Johns Hopkins University.

New ideas about atrial fibrillation 50 years on

Stanley Nattel

Montreal Heart Institute Research Center, Department of Medicine, Montreal Heart Institute and University of Montreal, Montreal, Quebec, H1T 1C8, Canada (e-mail: nattel@icm.umontreal.ca)

Atrial fibrillation is a condition in which control of heart rhythm is taken away from the normal sinus node pacemaker by rapid activity in different areas within the upper chambers (atria) of the heart. This results in rapid and irregular atrial activity and, instead of contracting, the atria only quiver. It is the most common cardiac rhythm disturbance and contributes substantially to cardiac morbidity and mortality. For over 50 years, the prevailing model of atrial fibrillation involved multiple simultaneous re-entrant waves, but in light of new discoveries this hypothesis is now undergoing re-evaluation.

Atrial fibrillation (AF) is characterized by rapid and irregular activation of the atrium, for example, 400–600 pulses of the atrium muscular wall per minute in humans. The occurrence of AF increases with age, with a prevalence rising from 0.5% of people in their 50s to nearly 10% of the octogenarian population^{1,2}. Several cardiac disorders predispose to AF, including coronary artery disease, pericarditis, mitral valve disease, congenital heart disease, congestive heart failure (CHF), thyrotoxic heart disease and hypertension. Many of these are thought to promote AF by increasing atrial pressure and/or by causing atrial dilation; however, the precise mechanistic links are incompletely defined. AF also occurs in individuals without any other evidence of heart or systemic disease — a condition known as ‘lone AF’.

Normally, heart rate is finely attuned to the body's metabolic needs through physiological control of the cardiac pacemaker function of the sinoatrial node (Fig. 1a), which maintains a rate of about 60 beats per minute at rest and can fire as rapidly as 180–200 times per minute at peak exercise. During AF, atrial cells fire at rates of 400–600 times per minute. If each atrial impulse were conducted to the ventricles, the extremely rapid ventricular rate would lead to ineffective cardiac contraction and rapid death. This is prevented by the filtering function of the atrioventricular (AV) node (Fig. 1b), which has a limited impulse-carrying capacity and through which atrial impulses must pass before activating the ventricles.

The ventricular rate during AF (the effective ‘heart rate’) is thus no longer under physiological control of the sinus node, but instead is determined by interaction between the atrial rate and the filtering function of the AV node. The ventricular rate during AF is typically in the region of 150 pulses per minute in the absence of drug therapy. In normal individuals, a brief period of AF may cause palpitations, chest discomfort and light-headedness. Sustained AF with an uncontrolled ventricular response rate can, by itself, cause severe CHF after several weeks to months, but this is reversible with proper rate and/or rhythm control³.

Owing to the loss of effective atrial contraction, and the irregular and excessively rapid ventricular rhythms that can be caused by AF, acute and sometimes life-threatening decompensation of otherwise compensated cardiac disease may occur. The loss of atrial contraction also leads to stasis of blood in the atria, which promotes clot formation and the occurrence of thromboemboli. These thromboemboli tend

to propagate, particularly to the brain but also to other organs (including the kidneys, mesenteric circulation and the heart itself), potentially leading to infarction. The thromboembolic risk is reduced by administration of oral anticoagulant drugs, but at the price of an increased risk of bleeding complications. These considerations probably account for the significant role of AF in the occurrence of stroke: AF is the single most important cause of ischaemic stroke in people older than 75 (ref. 4).

Cardiac arrhythmias have been treated traditionally with antiarrhythmic drugs that control the rhythm by altering cardiac electrical properties. However, the available drugs are not specific for atrial electrical activity and can have profound effects on ventricular electrophysiology. For example, K⁺-channel-blocking drugs that are used to treat AF can mimic potentially lethal congenital disorders of cardiac repolarization, as discussed elsewhere in this issue by Marban, pages 213–218.

It has become apparent over the past 15 years that the effects of antiarrhythmic drugs on the electrophysiology of the ventricles can themselves paradoxically lead to life-threatening rhythm disorders (so-called ‘proarrhythmia’) and increase mortality⁵. There has been, therefore, a shift towards non-pharmacological therapies for cardiac arrhythmias, including controlled destruction of arrhythmia-generating tissue (‘ablation therapy’) and implantable devices that can sense arrhythmias and terminate them with controlled electrical discharges. In contrast to a wide range of other cardiac arrhythmias, for which safe and highly effective non-pharmacological therapies have been developed, AF continues to be a challenge for both pharmacological and non-pharmacological approaches to treatment⁶, which has motivated a search for improved treatment modalities. One hope is that a better understanding of the fundamental mechanisms underlying AF will lead to safer and more effective mechanism-based therapeutic approaches.

Here I review the recent evolution of our understanding of the mechanisms of AF, highlighting experimental findings that have produced new insights and questions about the theory of the arrhythmia. In addition, I discuss the potential implications of this knowledge for therapeutic innovation.

Basic mechanisms of arrhythmia

To aid in the comprehension of AF mechanisms, the basic electrophysiological mechanisms of arrhythmia are presented in Fig. 2. Abnormal impulse formation can lead to

focal ectopic (that is, arising from an area other than the sinus node) arrhythmia generators (Fig. 2a,b). Figure 2a illustrates the mechanism of 'automaticity', which is the basis of cardiac pacemaker function. Various areas of the heart can show automaticity, but cardiac rhythmicity is normally controlled by the sinus node, which is intrinsically the fastest pacemaker.

Inward currents (positive ions entering the cell) depolarize the cell membrane (making the cell interior more positive), and outward currents repolarize it (making the cell interior more negative). Pacemaker activity results from a change in the balance of cardiac currents in the resting (diastolic) phase of the action potential, such that inward currents predominate and lead to progressive diastolic depolarization. When the membrane potential reaches a critical value, called the 'threshold potential', the cell fires (Fig. 2a, 1). If the slope of diastolic depolarization is increased in a region outside the sinus node (for example, by acute myocardial ischaemia; Fig. 2a, dashed line), the cell will reach threshold earlier and generate ectopic action potentials (Fig. 2a, 2) at a rapid rate.

Abnormal focal activity can also arise from 'afterdepolarizations' (Fig. 2b). As discussed in detail elsewhere in this issue (see review by Bers, pages 198–205), the free intracellular Ca^{2+} concentration rises sharply during the depolarized (systolic) phase of the action potential, and diastolic relaxation ensues when cytosolic Ca^{2+} is reduced by uptake into the sarcoplasmic reticulum and by extrusion from the cell through transmembrane $\text{Na}^+/\text{Ca}^{2+}$ exchange (NCX). The latter process is electrogenic, exchanging three Na^+ ions for each Ca^{2+} (one net positive charge moves in the direction of Na^+ transport in each cycle), and produces an inward current during Ca^{2+} extrusion. This inward current can depolarize the cell, generating an afterdepolarization (Fig. 2b, 2). If afterdepolarizations reach the threshold potential (Fig. 2b, 3), spontaneous action potentials will result, producing ectopic firing if they arise in tissues outside the sinus node. This form of afterdepolarization is favoured by

conditions that increase NCX current, such as excessive intracellular concentrations of Ca^{2+} and enhanced NCX activity.

Re-entry arises from abnormal impulse propagation between different zones of tissue (Fig. 2c, I and II). After initial depolarization of an action potential, Na^+ channels are inactivated and the cell cannot be re-fired until the cell repolarizes to a potential (about -60 mV) at which Na^+ channels recover from inactivation — a time referred to as the 'refractory period'. An ectopic complex (Fig. 2c, 2) arising in zone II during the refractory period of action potential 1 in zone I will initially fail to activate zone I, but may propagate through an alternative pathway to return to zone I when its refractory period is over, causing reactivation at this site (Fig. 2c, 3). The impulse will now leave zone I and move towards zone II and, if the time to return to zone II is sufficiently long, then zone II will be reactivated (Fig. 2c, 4).

If the conditions are correct (that is, there is an appropriate substrate for re-entry), zones I and II can repeatedly reactivate each other once re-activation has been initiated, resulting in persistent re-entrant activity. Re-entry can occur in a single circuit, producing rapid regular firing, or multiple unstable re-entry circuits can coexist simultaneously, producing more irregular (fibrillatory) activity. The

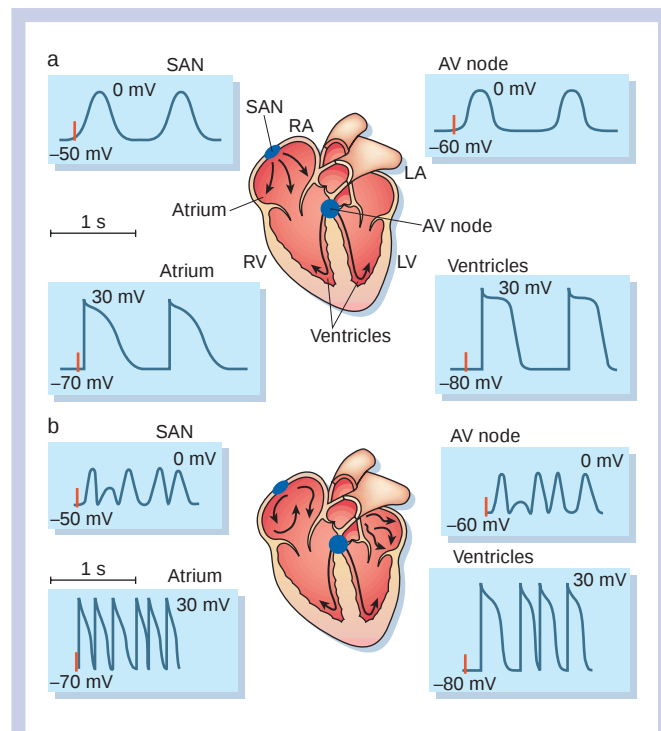


Figure 1 Diagram of electrical activity during atrial fibrillation. **a**, Normal rhythm; **b**, atrial fibrillation. Representative action potentials are shown from the sinoatrial node (SAN), atrium, AV node and ventricles. The vertical line on each action potential recording corresponds to a common time reference. LA, left atrium; LV, left ventricle; RA, right atrium; RV, right ventricle.

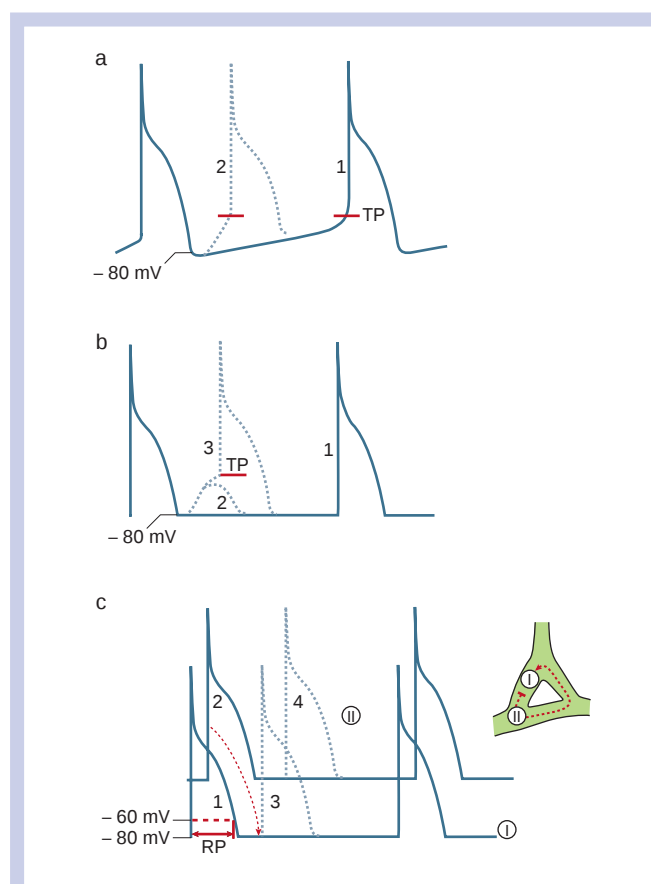


Figure 2 Cellular mechanisms of atrial arrhythmia generation. **a**, Normal 'automaticity' occurs when a spontaneously depolarizing cell reaches threshold potential and fires (1). When the rate of depolarization increases, abnormally rapid firing will occur (2). **b**, Afterdepolarizations are depolarizations caused by excessively large inward currents carried by the $\text{Na}^+/\text{Ca}^{2+}$ exchanger (2). If afterdepolarizations are large enough to reach threshold, premature ectopic action potentials result (3) before the next expected normal action potential (1). **c**, Re-entry occurring between two tissue zones, I and II, which are connected as shown on the right. A premature activation (2) in zone II, which fails to initiate firing in zone I because zone I is still refractory, may conduct back (red dashed line) to zone I at a time when it can respond with an action potential (3). This action potential may propagate to initiate (4) in zone II, and the process can continue indefinitely. RP, refractory period; TP, threshold potential.

refractory period is clearly a key determinant of re-entry — long refractory periods make it more likely that the circulating impulse will encounter tissue that is still refractory and will die out.

Classical mechanisms of atrial fibrillation

The conceptual framework for understanding AF mechanisms has been grounded in ideas developed in the early twentieth century⁷. The principal competing notions at the time were that AF is caused by rapidly discharging, spontaneously active, atrial ectopic foci (Fig. 3a), by a single re-entrant circuit (Fig. 3b), or by multiple functional re-entrant circuits (Fig. 3c). For multiple-circuit re-entry excitation (Fig. 3c), irregular atrial activity is a consequence of the primary arrhythmia mechanism. For the rapid focal and single-circuit mechanisms, irregularity is presumed to result from interactions between high-frequency wavefronts produced by the primary generator (the ectopic focus or primary re-entrant circuit) and the spatially variable refractory properties of atrial tissue ('fibrillatory conduction').

These mechanisms have significant implications for potential therapies. Multiple-circuit re-entry should be amenable to interventions that interfere with the ability of the re-entering circuits to perpetuate themselves. Such interventions include drugs that increase the refractory period and surgical division of the atria into electrically isolated areas, as exemplified by the 'Maze' procedure⁸. Ectopic mechanisms would be susceptible to drugs that suppress automaticity and to targeted destruction of ectopic foci by surgery or catheter-based approaches. Single-circuit re-entry should be suppressed by drugs that prolong the refractory period and inhibit re-entry, and by ablating key components of the re-entrant pathway.

Over the past 50 years, multiple-circuit re-entry has been the dominant conceptual model of AF. The ectopic focus and single-circuit concepts fell largely into disfavour. Particularly influential has been the work of Moe and co-workers⁹, who emphasized the role of multiple re-entrant wavelets in the perpetuation of AF. An important component of this theoretical framework is the concept of the 'wavelength of re-entry', as developed by Allessie and colleagues^{10,11}. The wavelength is the distance travelled by the electrical impulse in one refractory period, which is the product of the refractory period and the conduction velocity. If the pathlength of the potential circuit is smaller than the wavelength, the impulse will traverse the circuit and return to its starting point in a time shorter than the refractory period, forcing it to impinge on still-refractory tissue and die out. Thus, the wavelength is the shortest pathlength that can sustain re-entry.

According to the 'leading circle' hypothesis of Allessie *et al.* (ref. 10; and Fig. 4a), functional re-entry naturally establishes itself in a pathlength the size of the wavelength. The number of re-excitation

waves that can be accommodated is then a simple function of atrial size and the wavelength: decreased wavelength decreases the minimum circuit size, which increases the number of circuits that can be accommodated, which in turn favours multiple-circuit re-entry and tends to perpetuate AF (Fig. 4c). Traditionally, therefore, the primary approach to AF has been to increase the refractory period (and thereby the wavelength), which limits the number of functional circuits that can be maintained so that AF cannot sustain itself.

An observation incompatible with leading circle theory is the response of AF to antiarrhythmic drugs that block Na^+ channels. Such agents are effective in terminating AF, but according to leading circle theory should promote AF because they decrease conduction velocity and thereby decrease the wavelength.

Observations that challenge the classical viewpoint

Observations obtained over the past 5 years have challenged the previously prevailing notion that all AF is caused by multiple-circuit re-entry. Optical mapping studies of AF in sheep hearts point to a primary local generator, consisting of either a single small re-entrant circuit or an ectopic focus^{12,13}. Left atrial sources of activity seem to be particularly important^{12–14}. Left atrial predominance may be related to the location of the pulmonary veins in the left atrium, which seem to have a highly significant role (ref. 15; and see below), or to ionic differences that lead to shorter left atrial refractory periods that favour re-entry¹⁶. There is also evidence to show that single-circuit re-entry maintains AF in experimental CHF^{17,18}.

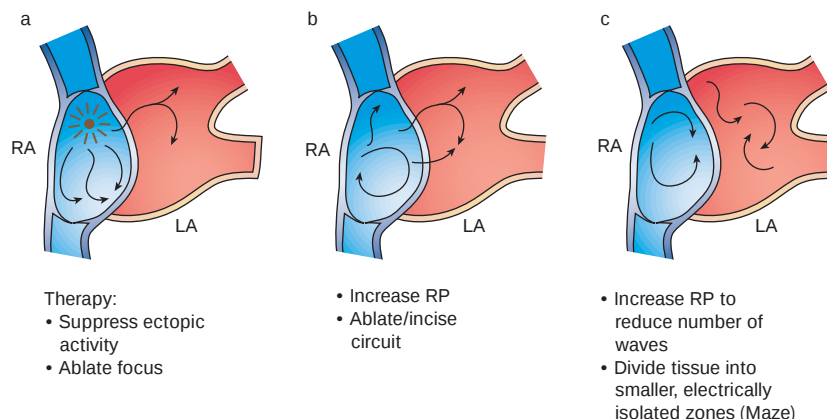
Thus, the recently evolving evidence has thrown us back to the debates of the early twentieth century, suggesting that ectopic activity, single-circuit re-entry and multiple circuit re-entry may all be involved in AF.

The electrophysiological basis of atrial fibrillation

Evolving clinical evidence shows that AF almost invariably occurs in a setting of atrial electrical dysfunction that provides a favourable substrate for the arrhythmia. Transmembrane ionic currents are key determinants of the arrhythmia mechanisms shown in Fig. 2. Figure 5a shows a representative human atrial action potential. Inward (depolarizing) currents are indicated by a downward arrow and outward (repolarizing) currents by an upward arrow. I_{K1} is the background current responsible for the considerable resting K^+ conductance that sets the resting potential to between -70 and -80 mV.

Cell firing is caused by rapid depolarization through a large Na^+ current (I_{Na}) that brings the cell from its resting potential to a value in the region of $+40$ mV, providing the electrical energy for cardiac conduction. The cell then partially repolarizes through a transient

Figure 3 Conceptual models of atrial fibrillation in the early twentieth century, along with therapeutic implications. Ectopic foci (a) and single re-entrant circuits (b) generating atrial fibrillation can be located in either atrium, but are shown here as arising in the right atrium. In c, atrial fibrillation is maintained by multiple re-entry circuits with spatial and temporal variability. LA, left atrium; RA, right atrium; RP, refractory period.



outward K^+ current (I_{to}), inactivation of which produces a notch in the action potential. This is followed by a relatively flat portion of the action potential (the 'plateau'), which is maintained by an inward L-type Ca^{2+} current (I_{Ca}). A series of K^+ currents that activate in a time-dependent way and show little inactivation — the so-called 'delayed rectifiers' (I_K) — leads to cellular repolarization. In human atrium, I_K has three components: an 'ultra-rapid' component (I_{Kur}), a 'rapid' component (I_{Kr}) and a 'slow' component (I_{Ks}). Spontaneously automatic cells are depolarized by an inward pacemaker current (I_f). NCX also carries an inward current during terminal repolarization and for a short time thereafter.

The balance between plateau inward and outward currents determines the action potential duration (APD): increased inward current prolongs the action potential, and increased outward current abbreviates it. APD governs the time from cellular depolarization to recovery of excitability at about -60 mV; the ionic current balance therefore determines the refractory period and the likelihood of re-entry. Alterations in ionic currents that increase APD and thereby the refractory period can be used to prevent AF. For example, many clinically used drugs prolong APD and refractoriness by inhibiting I_{Kr} . They are effective in preventing AF, but can produce dangerous ventricular arrhythmias by interfering with ventricular repolarization (see review by Marban, pages 213–218). I_{Ks} and I_{Kur} are under strong adrenergic control¹⁹, and their stimulation might contribute to AF that occurs in situations of increased adrenergic tone. I_{Kur} is carried by Kv1.5 channels that are expressed functionally in human atrium but not ventricle — inhibiting these channels may provide a means of preventing AF without the risk of ventricular proarrhythmia^{20–22}.

An important advance in our understanding of AF was made with the recognition that AF, once initiated, alters atrial electrophysiological properties in a manner that favours the ease of inducing and maintaining the arrhythmia — a process called 'electrical remodeling'²³. The principal mechanisms involved are shown in Fig. 6. The roughly tenfold atrial rate increase caused by AF is the primary stimulus to remodeling, and similar changes are produced by any form of sufficiently rapid atrial tachycardia²⁴. Ca^{2+} enters the cells through I_{Ca} with each action potential, so a tenfold increase in atrial rate substantially increases cellular Ca^{2+} loading²⁵.

Progressive Ca^{2+} loading threatens cell viability, and the cells respond to minimize the impact of increased rate on intracellular Ca^{2+} load. Short-term defence mechanisms include voltage-dependent and intracellular Ca^{2+} -concentration-dependent inactivation of I_{Ca} (ref. 26). Over the longer term, the concentration of messenger RNA encoding the pore-forming I_{Ca} α -subunit^{27–29} decreases, which in turn decreases I_{Ca} (refs 30–32). Both short- and long-term decreases in I_{Ca} reduce Ca^{2+} entry and help to prevent Ca^{2+} overload; however, because I_{Ca} is a key contributor to the action potential plateau (Fig. 5), reduced I_{Ca} decreases APD, reduces the refractory period, and promotes the induction and maintenance of AF by multiple-circuit re-entry³³. AF that begins by any mechanism causes electrical remodeling, which by promoting multiple-circuit re-entry will make this a 'final common pathway' of AF irrespective of the initial mechanism (Fig. 7).

In addition to downregulating I_{Ca} , AF induces many other changes, consistent with a substantial cellular insult caused by excessively rapid activation. Cellular Ca^{2+} handling is altered, decreasing the release of systolic Ca^{2+} (ref. 34) in association with altered concentrations of intracellular Ca^{2+} -handling proteins^{29,35}. Cellular myolysis occurs, along with changes that suggest a return to a more fetal phenotype³⁶. Decreased release of systolic Ca^{2+} and myolysis impair atrial contractility, contributing to the occurrence of atrial blood stasis and thromboembolic events after termination of AF³⁷. I_{Na} seems to be decreased, possibly contributing to a slowly developing decrease in atrial conduction that may help to promote AF^{33,38}. I_{to} is decreased^{30,32,39} and I_{K1} may be altered^{32,39}; however, the physiological significance of these changes in K^+ current are currently unclear. Finally, AF has been associated with altered expression of connexin

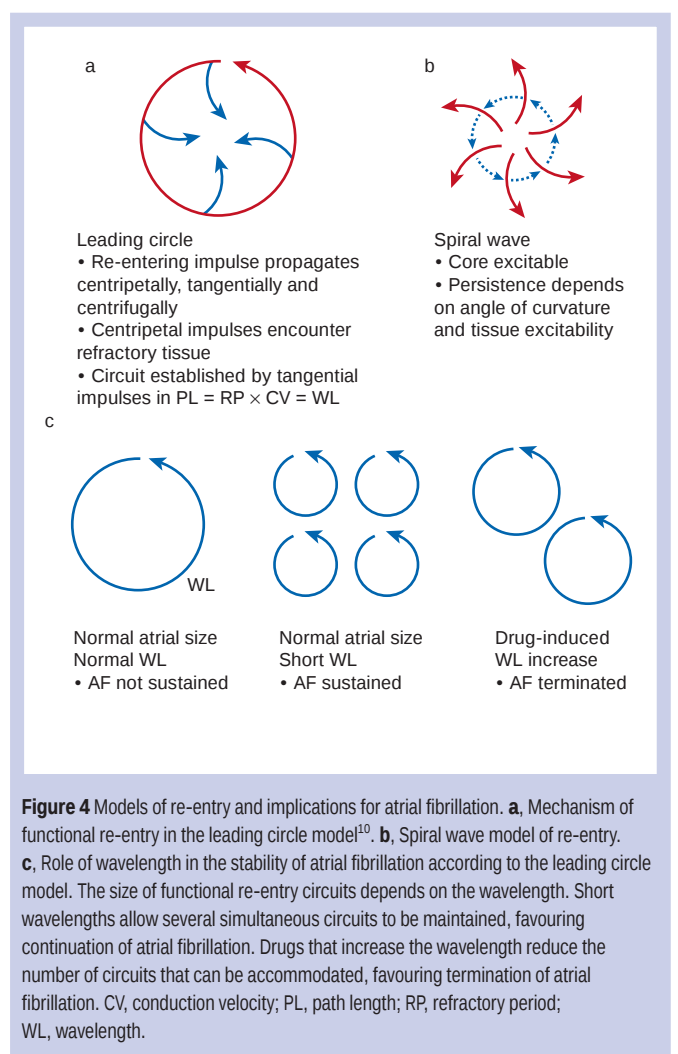


Figure 4 Models of re-entry and implications for atrial fibrillation. **a**, Mechanism of functional re-entry in the leading circle model¹⁰. **b**, Spiral wave model of re-entry. **c**, Role of wavelength in the stability of atrial fibrillation according to the leading circle model. The size of functional re-entry circuits depends on the wavelength. Short wavelengths allow several simultaneous circuits to be maintained, favouring continuation of atrial fibrillation. Drugs that increase the wavelength reduce the number of circuits that can be accommodated, favouring termination of atrial fibrillation. CV, conduction velocity; PL, path length; RP, refractory period; WL, wavelength.

channel proteins that govern intercellular electrical communication^{40–42}. There is evidence for spatially heterogeneous reductions in expression of connexin 40 (ref. 41), but the results of available studies have been inconsistent^{40–42} and the significance of these changes in connexin remains unclear.

Experimental CHF also promotes AF and produces atrial ionic remodeling⁴³. In CHF, inward I_{Ca} is reduced much less ($\sim 30\%$) than in atrial tachycardia remodeling ($\sim 65\%$). Outward I_{Ks} , which is unaffected by atrial tachycardia, is decreased by about 30% and inward NCX is increased⁴³. Thus, unlike atrial-tachycardia-induced ionic remodeling, CHF-induced remodeling does not reduce APD and does not in itself favour re-entry. By contrast, CHF changes the architecture of atrial tissue, causing a substantial increase in fibrous tissue content (fibrosis) within and between muscle bundles, which interferes with electrical conduction and causes AF that often seems to be due to single-circuit re-entry^{17,18,44}. In addition, the increased NCX promotes afterdepolarization-related atrial ectopic firing and AF⁴⁵. Figure 5 summarizes the ionic determinants of AF, including both intrinsic and extrinsic determinants.

Molecular and genetic factors in atrial fibrillation

Although electrical remodeling accounts for the self-promoting nature of AF once it has begun, other factors must lead to the initial occurrence of AF. Individuals affected with AF are known to have upregulated levels of atrial extracellular signal-related kinase (ERK) and angiotensin-converting enzyme (ACE)⁴⁶, but a decreased density of angiotensin-II type 1 receptors and an increased density of angiotensin-II type 2 receptors⁴⁷. Carboxypeptidase activity and

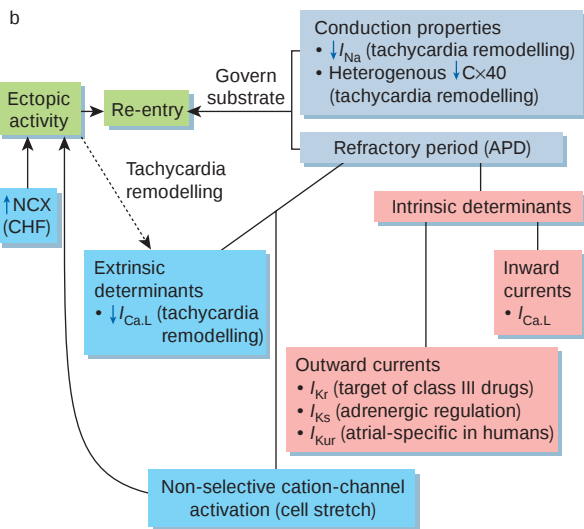
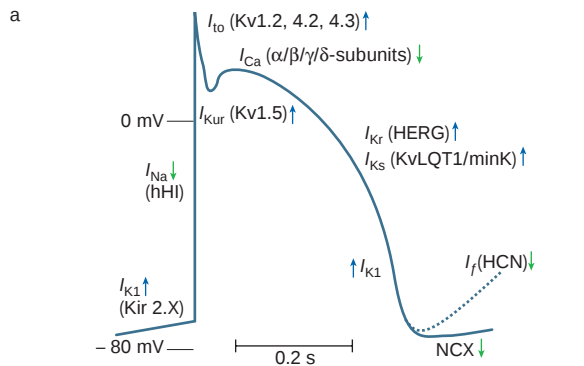


Figure 5 Ionic determinants of atrial fibrillation. **a**, Action potential showing the principal currents that flow in each phase, with the corresponding subunit clones shown in parentheses. **b**, The substrate for re-entry is governed by the refractory period and the conduction velocity. The refractory period depends on the action potential duration (APD), which is governed by the intrinsic ionic current properties of atrial tissue. Conduction properties are determined by the function of Na^+ channels and connexins. Reductions in outward currents increase the refractory period, prevent re-entry and oppose atrial fibrillation. Reductions in inward current and increases in outward current reduce the refractory period and promote atrial fibrillation. Increased $\text{Na}^+/\text{Ca}^{2+}$ exchange promotes afterdepolarization-related ectopic activity, which can trigger re-entry in the presence of an appropriate substrate or cause atrial tachycardias that induce electrical remodelling and produce a re-entry substrate. Activation of stretch-induced channels can promote both ectopic activity and re-entry. Cx40, connexin 40; CHF, congestive heart failure.

protein levels are also decreased in chronic AF⁴⁸. Atrial tissue from AF-affected individuals shows both nuclei that are positive for TUNEL (terminal deoxynucleotidyl transferase-mediated dUTP nick end labelling) and increased concentrations of the apoptotic enzyme caspase-3, which suggests that apoptosis is occurring⁴⁹. In experimental CHF, the promotion of AF is preceded by increased atrial concentrations of angiotensin-II and by activation of mitogen-activated protein kinases including ERK⁵⁰. A rise in atrial caspase-3 concentration, associated with infiltration of white blood cells, apoptosis and cell death, is seen early in experimental CHF and is followed by progressive fibrosis⁵¹.

These results suggest that, in CHF, atrial cell death pathways are activated, followed by replacement fibrosis, which promotes intra-

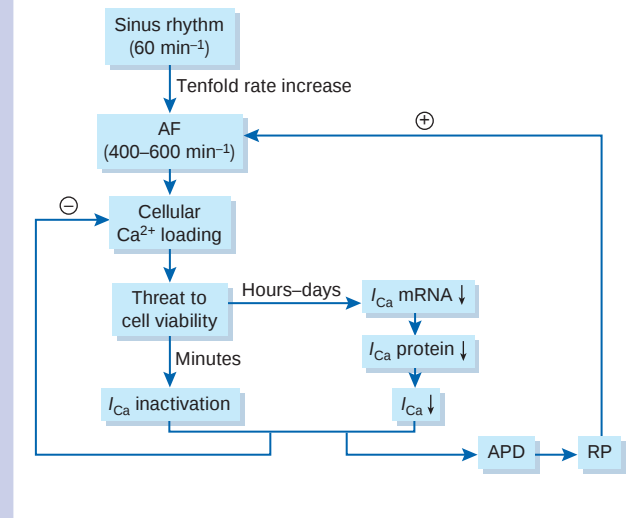


Figure 6 Changes in cell Ca^{2+} loading caused by atrial fibrillation and consequent adaptive responses. Adaptations to atrial tachycardia decrease Ca^{2+} loading by reducing I_{Ca} , but at the expense of a shorter action potential duration (APD) and a greater propensity for atrial fibrillation. RP, refractory period.

atrial re-entry. Inhibition of ACE reduces CHF-induced activation of ERK and fibrosis, and also decreases the AF-promoting effects of CHF⁵⁰. The effectiveness of ACE inhibition shows that interrupting signalling pathways can prevent development of the AF substrate—a process that might potentially be used in new therapeutic approaches. A recent study supporting this notion showed that ACE inhibition reduces the incidence of AF after myocardial infarction in people with left ventricular dysfunction⁵².

Atrial fibrillation can occur on a familial basis, pointing to a genetic cause of the arrhythmia in some individuals⁵³. Linkage analyses have identified possible loci on chromosome 10 in two kindreds⁵⁴. Identifying the gene of susceptibility, coupled with defining the sequence and function of the protein that it encodes, has the potential to provide both insight into the pathophysiology of the arrhythmia and diagnostic tools with which to identify susceptible individuals.

Studies of transgenic mouse models of AF are still in their infancy but promise to advance our understanding of the role of defined genes in the pathophysiology of the arrhythmia. Mice with a homozygous deficiency in the gene encoding connexin 40 have slowed intra-atrial and AV nodal conduction, along with inducible atrial tachyarrhythmias⁵⁵. Mice overexpressing cardiac RhoA, a low molecular-weight GTPase, develop atrial fibrillation and AV block, and show progressive deterioration of ventricular function⁵⁶. Mice with cardiac overexpression of a constitutively activated form of transforming growth factor- β have atrial but not ventricular fibrosis⁵⁷. In addition, preliminary data suggest that these mice are susceptible to AF⁵⁸, compatible with the notion that atrial fibrosis is an important AF-promoting factor.

The molecular events leading to ionic remodelling remain incompletely understood. On the basis of evidence indicating that Ca^{2+} overload has a central role in AF, the efficacy of I_{Ca} blockers in atrial tachycardia remodelling has been studied. L-type Ca^{2+} -channel blockers are effective in short-term remodelling^{59–61}, but are ineffective in remodelling caused by longer periods of atrial tachycardia^{62,63}. Mibefradil, a drug that blocks both L-type Ca^{2+} channels and the high-threshold T-type Ca^{2+} channels, prevents atrial tachycardia remodelling effectively, but it remains uncertain whether this action is due to inhibition of I_{Ca} or to collateral drug actions^{63,64}. The T-type Ca^{2+} current has been difficult to demonstrate in human atrial myocytes⁶⁵. Recent work has suggested that oxidant stress may be



normal and abnormal cardiac pacemaker function, and may provide important knowledge regarding the pathophysiology of AF and other cardiac arrhythmias.

Theoretical aspects

The development of 'spiral wave' theory has provided a new model for the fundamental properties of cardiac re-entry (ref. 77; and Fig. 4b). According to this theory, re-entry is maintained through the ability of circulating spiral waves to perpetuate in media with sufficient excitability to support the angle of spiral curvature. Evidence has been provided for a non-activated, excitable core at the centre of re-entry circuits during AF⁷⁸ — as predicted by spiral wave theory — in contrast to the constantly excited and refractory tissue predicted by leading circle theory (Fig. 2a).

Recent work using a two-dimensional computer model of AF that is based on realistic atrial cellular ionic properties suggests that Na⁺ blockade may terminate AF by reducing excitability, thereby causing the spiral wave activity that maintains AF to die out despite a decrease in the wavelength⁷⁹. This theoretical work can account for recent experimental observations of the action of class I antiarrhythmic drugs in AF⁸⁰.

Synthesis and future directions

The sole, multiple-circuit re-entry mechanism for AF, which was the predominant notion for about 50 years until the mid-1990s, has been challenged by more recent work showing the complexity of AF mechanisms and the validity of competing concepts proposed in the early twentieth century. Furthermore, the inter-relationships among AF mechanisms and the determinants of their occurrence have been highlighted (Fig. 7), raising several new questions. What are the molecular signals leading from atrial tachycardia to electrical remodelling? By what signalling mechanisms does CHF lead to atrial fibrosis and atrial ionic remodelling? What are the mechanisms of AF associated with hypertension, coronary artery disease, thyrotoxicosis and senescence? Why do the pulmonary veins seem to be so important in initiating, and possibly maintaining, AF? How does atrial tachycardia affect pulmonary vein activity? What are the precise, relative roles of ectopic activity, single-circuit re-entry and multiple-circuit re-entry in maintaining clinical AF? What are the genes responsible for familial AF, and how do they lead to the arrhythmia? Can improved knowledge of the dynamical basis of AF lead to pacing modalities that can prevent or stop the arrhythmia? Can interventions against remodelling safely and effectively prevent the development of the substrate for AF?

In the near future, the answers to these and related questions are likely to increase our understanding of the mechanisms underlying AF and to lead to new and improved possibilities in prevention and therapy. □

- Benjamin, E. J. *et al.* Impact of atrial fibrillation on the risk of death. The Framingham heart study. *Circulation* **98**, 946–952 (1998).
- Ho, K. K., Pinsky, J. L., Kannel, W. B. & Levy, D. The epidemiology of heart failure: the Framingham study. *J. Am. Coll. Cardiol.* **22**, 6A–13A (1993).
- Fenelon, G., Wijns, W., Andries, E. & Brugada, P. Tachycardiomyopathy: mechanisms and clinical implications. *Pacing Clin. Electrophysiol.* **19**, 95–106 (1996).
- Hart, R. G. & Halperin, J. L. Atrial fibrillation and stroke: concepts and controversies. *Stroke* **32**, 803–808 (2001).
- Nattel, S. Experimental evidence for proarrhythmic mechanisms of antiarrhythmic drugs. *Cardiovasc. Res.* **37**, 567–577 (1998).
- Nattel, S. Newer developments in the management of atrial fibrillation. *Am. Heart J.* **130**, 1094–1106 (1995).
- Garrey, W. E. Auricular fibrillation. *Physiol. Rev.* **4**, 215–250 (1924).
- Cox, J. L. & Ad, N. New surgical and catheter-based modifications of the Maze procedure. *Semin. Thorac. Cardiovasc. Surg.* **12**, 68–73 (2000).
- Moe, G. K., Rheinboldt, W. C. & Abildskov, J. A. A computer model of atrial fibrillation. *Am. Heart J.* **67**, 200–220 (1964).
- Allessie, M. A., Bonke, F. I. & Schopman, F. J. Circus movement in rabbit atrial muscle as a mechanism of tachycardia. III. The "leading circle" concept: a new model of circus movement in cardiac tissue without the involvement of an anatomical obstacle. *Circ. Res.* **41**, 9–18 (1977).
- Rensma, P. L., Allessie, M. A., Lammers, W. J., Bonke, F. I. & Schalij, M. J. Length of excitation wave and susceptibility to reentrant atrial arrhythmias in normal conscious dogs. *Circ. Res.* **62**, 395–410 (1988).
- Mandapati, R., Skanes, A., Chen, J., Berenfeld, O. & Jalife, J. Stable microreentrant sources as a mechanism of atrial fibrillation in the isolated sheep heart. *Circulation* **101**, 194–199 (2000).
- Mansour, M. *et al.* Left-to-right gradient of atrial frequencies during acute atrial fibrillation in the isolated sheep heart. *Circulation* **103**, 2631–2636 (2001).
- Morillo, C. A., Klein, G. J., Jones, D. L., Guiraudon, C. M. Chronic rapid atrial pacing. Structural, functional, and electrophysiological characteristics of a new model of sustained atrial fibrillation. *Circulation* **91**, 1588–1595 (1995).
- Haissaguerre, M. *et al.* Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. *N. Engl. J. Med.* **339**, 659–666 (1998).
- Li, D., Zhang, L., Kneller, J., Shi, H. & Nattel, S. Ionic mechanism of repolarization differences between canine right and left atrium. *Circ. Res.* **88**, 1168–1175 (2001).
- Derakhchan, K. *et al.* Method for simultaneous epicardial and endocardial mapping of the in vivo canine heart: application to atrial conduction properties and arrhythmia mechanisms. *J. Cardiovasc. Electrophysiol.* **12**, 542–555 (2001).
- Nattel, S., Li, D. & Yue, L. Basic mechanisms of atrial fibrillation—very new insights into very old ideas. *Annu. Rev. Physiol.* **62**, 51–77 (2000).
- Li, G.-R., Feng, J., Wang, Z., Fermini, B. & Nattel, S. Adrenergic modulation of ultrarapid delayed rectifier K⁺ current in human atrial myocytes. *Circ. Res.* **78**, 903–915 (1996).
- Wang, Z., Fermini, B. & Nattel, S. Sustained depolarization-induced outward current in human atrial myocytes: evidence for a novel delayed rectifier K⁺ current similar to Kv1.5 cloned channel currents. *Circ. Res.* **73**, 1061–1076 (1993).
- Li, G.-R., Feng, J., Yue, L., Carrier, M. & Nattel, S. Evidence for two components of delayed rectifier potassium current in human ventricular myocytes. *Circ. Res.* **78**, 689–696 (1996).
- Feng, J., Wible, B., Li, G.-R., Wang, Z. & Nattel, S. Antisense oligonucleotides directed against Kv1.5 mRNA specifically inhibit ultrarapid delayed rectifier potassium current in cultured adult human atrial myocytes. *Circ. Res.* **80**, 572–579 (1997).
- Wijffels, M. C. *et al.* Atrial fibrillation begets atrial fibrillation. A study in awake chronically instrumented goats. *Circulation* **92**, 1954–1968 (1995).
- Wijffels, M. C., Kirchhof, C. J., Dorland, R., Power, J. & Allessie, M. A. Electrical remodeling due to atrial fibrillation in chronically instrumented conscious goats: roles of neurohumoral changes, ischemia, atrial stretch, and high rate of electrical activation. *Circulation* **96**, 3710–3720 (1997).
- Sun, H., Chartier, D., Leblanc, N. & Nattel, S. Intracellular calcium changes and tachycardia-induced contractile dysfunction in canine atrial myocytes. *Cardiovasc. Res.* **49**, 751–761 (2001).
- Courtemanche, M., Ramirez, R. F. & Nattel, S. Ionic mechanisms underlying human atrial action potential properties: insights from a mathematical model. *Am. J. Physiol.* **275**, H301–H321 (*Heart Circ. Physiol.* 44) (1998).
- Yue, L. *et al.* Molecular mechanisms underlying ionic remodeling in a dog model of atrial fibrillation. *Circ. Res.* **84**, 776–784 (1999).
- Brundel, B. J. *et al.* Gene expression of proteins influencing the calcium homeostasis in patients with persistent and paroxysmal atrial fibrillation. *Cardiovasc. Res.* **42**, 443–454 (1999).
- Lai, L.-P. *et al.* Down-regulation of L-type calcium channel and sarcoplasmic reticular Ca²⁺-ATPase mRNA in human atrial fibrillation without significant change in the mRNA of ryanodine receptor, claquestrin and phospholamban. An insight into the mechanism of atrial electrical remodeling. *J. Am. Coll. Cardiol.* **33**, 1231–1237 (1999).
- Yue, L. *et al.* Ionic remodeling underlying action potential changes in a canine model of atrial fibrillation. *Circ. Res.* **81**, 512–525 (1997).
- Van Wagoner, D. R. *et al.* Atrial L-type Ca²⁺ currents and human atrial fibrillation. *Circ. Res.* **85**, 428–436 (1999).
- Bosch, R. F. *et al.* Ionic mechanisms of electrical remodeling in human atrial fibrillation. *Cardiovasc. Res.* **44**, 121–131 (1999).
- Gaspo, R., Bosch, R. F., Talajic, M. & Nattel, S. Functional mechanisms underlying tachycardia-induced sustained atrial fibrillation in a chronic dog model. *Circulation* **96**, 4027–4035 (1997).
- Sun, H., Gaspo, R., Leblanc, N. & Nattel, S. The cellular mechanisms of atrial contractile dysfunction caused by sustained atrial tachycardia. *Circulation* **98**, 719–727 (1998).
- Ohkusa, T. *et al.* Alterations in cardiac sarcoplasmic reticulum Ca²⁺ regulatory proteins in the atrial tissue of patients with chronic atrial fibrillation. *J. Am. Coll. Cardiol.* **34**, 255–263 (1999).
- Ausma, J. *et al.* Structural changes of atrial myocardium due to sustained atrial fibrillation in the goat. *Circulation* **96**, 3157–3163 (1997).
- Black, I. W. *et al.* Exclusion of atrial thrombus by transesophageal echocardiography does not preclude embolism after cardioversion of atrial fibrillation. A multicenter study. *Circulation* **89**, 2509–2513 (1994).
- Gaspo, R., Bosch, R. F., Bou-Aboud, E. & Nattel, S. Tachycardia-induced changes in sodium current in a chronic dog model of atrial fibrillation. *Circ. Res.* **81**, 1045–1052 (1997).
- Van Wagoner, D. R., Pond, A. L., McCarthy, P. M., Trimmer, J. S. & Nerbonne, J. M. Outward K⁺ current densities and Kv1.5 expression are reduced in chronic human atrial fibrillation. *Circ. Res.* **80**, 772–781 (1997).
- Elvan, A., Huang, X. D., Pressler, M. L. & Zipes, D. P. Radiofrequency catheter ablation of the atrial eliminates pacing-induced sustained atrial fibrillation and reduces connexin43 in dogs. *Circulation* **96**, 1675–1685 (1997).
- van der Welden, H. M. W. *et al.* Gap junctional remodeling in relation to stabilization of atrial fibrillation in the goat. *Cardiovasc. Res.* **46**, 476–486 (2000).
- Dupont, E. *et al.* The gap-junctional protein connexin40 is elevated in patients susceptible to postoperative atrial fibrillation. *Circulation* **103**, 842–849 (2001).
- Li, D. *et al.* The effects of experimental heart failure on atrial cellular and ionic electrophysiology. *Circulation* **101**, 2631–2638 (2000).
- Li, D., Fareh, S., Leung, T. K. & Nattel, S. Promotion of atrial fibrillation by heart failure in dogs: atrial remodeling of a different sort. *Circulation* **100**, 87–95 (1999).
- Fenelon, G., Manders, T. & Stamler, B. S. Atrial tachycardia in dogs with ventricular pacing-induced congestive heart failure originates from multiple foci in the crista terminalis and pulmonary veins: experimental evidence supporting the "atrial ring of fire" hypothesis. *Circulation* **96**, I-237 (1999).
- Goette, A. *et al.* Increased expression of extracellular signal-regulated kinase and angiotensin-converting enzyme in human atria during atrial fibrillation. *J. Am. Coll. Cardiol.* **35**, 1669–1677 (2000).
- Goette, A. *et al.* Regulation of angiotensin II receptor subtypes during atrial fibrillation in humans. *Circulation* **101**, 2678–2681 (2000).
- Lendeckel, U. *et al.* Expression and activity of ectopeptidases in fibrillating human atria. *J. Mol. Cell. Cardiol.* **33**, 1273–1281 (2001).

49. Aime-Sempe, C. *et al.* Myocardial cell death in fibrillating and dilated human right atria. *J. Am. Coll. Cardiol.* **34**, 1577–1586 (1999).
50. Li, D. *et al.* Effects of angiotensin converting enzyme inhibition on the development of the atrial fibrillation substrate in dogs with ventricular tachypacing-induced congestive heart failure. *Circulation* **104**, 2608–2614 (2001).
51. Cardin, S., Li, D., Thorin, E., Leung, T. K. & Nattel, S. Role of apoptosis and tissue fibrosis in arrhythmogenic atrial structural remodeling in a canine model of congestive heart failure (abstract). *Circulation* **104**, II-77 (2001).
52. Pedersen, O. D., Bagger, H., Kober, L. & Torp-Pedersen, C. Trandolapril reduces the incidence of atrial fibrillation after acute myocardial infarction in patients with left ventricular dysfunction. *Circulation* **100**, 376–380 (1999).
53. Beyer, F., Paul, T., Luhmer, I., Bertram, H. & Kallfelz, H. C. Familiäres idiopathisches Vorhofflimmern mit Bradyarrhythmie. *Z. Kardiol.* **82**, 674–677 (1993).
54. Brugada, R. *et al.* Identification of a genetic locus for familial atrial fibrillation. *N. Engl. J. Med.* **336**, 905–911 (1997).
55. Hagendorff, A. *et al.* Conduction disturbances and increased atrial vulnerability in connexin 40-deficient mice analyzed by transesophageal stimulation. *Circulation* **99**, 1508–1515 (1999).
56. Sah, V. P. *et al.* Cardiac-specific overexpression of RhoA results in sinus and atrioventricular nodal dysfunction and contractile failure. *J. Clin. Invest.* **103**, 1627–1634 (1999).
57. Nakajima, H. *et al.* Atrial but not ventricular fibrosis in mice expressing a mutant transforming growth factor- β 1 transgene in the heart. *Circ. Res.* **86**, 571–579 (2000).
58. Raesdama, A. *et al.* Inducibility of atrial arrhythmias in transgenic mice with selective atrial fibrosis due to overexpression of TGF β -1. *Pacing Clin. Electrophysiol.* **24**, 549 (2001).
59. Yu, W. C. *et al.* Tachycardia-induced change of atrial refractory period in humans: rate dependency and effects of antiarrhythmic drugs. *Circulation* **97**, 2331–2337 (1998).
60. Daoud, E. G. *et al.* Effect of verapamil and procainamide on atrial fibrillation-induced electrical remodeling in humans. *Circulation* **96**, 1542–1550 (1997).
61. Tieleman, R. G. *et al.* Verapamil reduces tachycardia-induced electrical remodeling of the atria. *Circulation* **95**, 1945–1953 (1997).
62. Lee, S. H. *et al.* Effect of verapamil on long-term tachycardia-induced atrial electrical remodeling. *Circulation* **101**, 200–206 (2000).
63. Fareh, S., Bénédeau, A. & Nattel, S. Differential efficacy of L- and T-type calcium channel blockers in preventing tachycardia-induced atrial remodeling in dogs. *Cardiovasc. Res.* **49**, 762–770 (2001).
64. Fareh, S., Bénédeau, A., Thibault, B. & Nattel, S. The T-type Ca^{2+} channel blocker mibefradil prevents the development of a substrate for atrial fibrillation by tachycardia-induced atrial remodeling in dogs. *Circulation* **100**, 2191–2197 (1999).
65. Li, G.-R. & Nattel, S. Properties of transmembrane Ca^{2+} current at physiologic temperatures relevant to the action potential in human atrial myocytes. *Am. J. Physiol.* **272**, H227–H235 (*Heart Circ. Physiol.* 41) (1997).
66. Carnes, C. A. *et al.* Ascorbate attenuates atrial pacing-induced peroxynitrite formation and electrical remodeling and decreases the incidence of postoperative atrial fibrillation. *Circ. Res.* **89**, e32–e38 (2001).
67. Hobbs, W. J., Van Gelder, I. C., Fitzpatrick, A. P., Crijns, H. J. & Garratt, C. J. The role of atrial electrical remodeling in the progression of focal atrial ectopy to persistent atrial fibrillation. *J. Cardiovasc. Electrophysiol.* **10**, 866–870 (1999).
68. Doshi, R. N. *et al.* Relation between ligament of Marshall and adrenergic atrial tachyarrhythmia. *Circulation* **100**, 876–883 (1999).
69. Tsai, C. F. *et al.* Initiation of atrial fibrillation by ectopic beats originating from the superior vena cava: electrophysiological characteristics and results of radiofrequency ablation. *Circulation* **102**, 67–74 (2000).
70. Wu, T. J. *et al.* Pulmonary veins and ligament of Marshall as sources of rapid activations in a canine model of sustained atrial fibrillation. *Circulation* **103**, 1157–1163 (2001).
71. Kumagai, K. *et al.* Role of rapid focal activation in the maintenance of atrial fibrillation originating from the pulmonary veins. *Pacing Clin. Electrophysiol.* **23**, 1823–1827 (2000).
72. Sueda, T. *et al.* Efficacy of pulmonary vein isolation for the elimination of chronic atrial fibrillation in cardiac valvular surgery. *Ann. Thorac. Surg.* **71**, 1189–1193 (2001).
73. Cheung, D. W. Electrical activity of the pulmonary vein and its interaction with the right atrium in the guinea-pig. *J. Physiol.* **314**, 445–456 (1981).
74. Chen, Y.-J., Chen, S.-A., Chang, M.-S. & Lin, C.-I. Arrhythmogenic activity of cardiac muscle in pulmonary veins of the dog: implication for the genesis of atrial fibrillation. *Cardiovasc. Res.* **48**, 265–273 (2000).
75. Bode, F. *et al.* Tarantula peptide inhibits atrial fibrillation. *Nature* **409**, 35–36 (2001).
76. Shi, W. *et al.* Distribution and prevalence of hyperpolarization-activated cation channel (HCN) mRNA expression in cardiac tissues. *Circ. Res.* **85**, e1–e6 (1999).
77. Beaumont, J., Davidenko, N., Davidenko, J. M. & Jalife, J. Spiral waves in two-dimensional models of ventricular muscle: formation of a stationary core. *Biophys. J.* **75**, 1–14 (1998).
78. Athill, C. A. *et al.* Transmembrane potential properties at the core of functional reentrant wave fronts in isolated canine right atria. *Circulation* **98**, 1556–1567 (1998).
79. Kneller, J. & Nattel, S. How do class I antiarrhythmic drugs terminate atrial fibrillation? A quantitative analysis based on a realistic ionic model (abstract). *Circulation* **104**, II-5 (2001).
80. Wijffels, M. C. *et al.* Widening of the excitable gap during pharmacological cardioversion of atrial fibrillation in the goat: effects of cibenzoline, hydroquinidine, flecainide, and D-sotalol. *Circulation* **102**, 260–267 (2000).

The failing heart

J. A. Towbin*† & N. E. Bowles

Departments of Pediatrics (Cardiology), *Molecular and Human Genetics, and †Cardiovascular Sciences, Baylor College of Medicine, Houston, Texas 77030, USA (e-mail: jtowbin@bcm.tmc.edu)

Cardiomyopathies are disorders affecting heart muscle that usually result in inadequate pumping of the heart. They are the most common cause of heart failure and each year kill more than 10,000 people in the United States. In recent years, there have been breakthroughs in understanding the molecular mechanisms involved in this group of conditions, with knowledge of the genetic basis for cardiomyopathies perhaps seeing the largest advance, enabling clinicians to devise improved diagnostic strategies and preparing the stage for new therapies.

Congestive heart failure (CHF) resulting from cardiomyopathy is a serious malady and a principal cause of death and disability in children and adults. This disorder is the most common disease that leads to heart transplant, with an associated healthcare cost in the United States of roughly US\$200 million per year¹. Cardiomyopathies are classified into four forms: dilated, hypertrophic, restrictive and arrhythmogenic right ventricular dysplasia/cardiomyopathy. Restrictive cardiomyopathy is the rarest form, accounting for only 5% of the total², but has the worst prognosis and poorest therapeutic options of all the cardiomyopathies³. Arrhythmogenic right ventricular dysplasia is a complex arrhythmogenic disorder associated with cardiomyopathy and characterized by gradual loss of myocytes and replacement by fatty and fibrous tissue⁴. In Italy the prevalence has been reported as 1:5,000 people, accounting for 20% of sudden deaths in young adults⁵ and 25% of cardiac sudden deaths among athletes⁶; the incidence and prevalence are unknown in the United States. In this review we will discuss the identification of genes responsible for cardiomyopathies, as well as some of the animal models now being developed to further characterize the pathogenesis of these disorders.

Dilated cardiomyopathy

Dilated cardiomyopathy (DCM) is the most common cause of CHF, affecting 40 people in every 100,000 of the

population⁷. Depending on the diagnostic criteria used, the annual incidence varies from 5 to 8 cases in a population of 100,000 (refs 7–9); the true incidence is probably underestimated by these numbers, as many asymptomatic cases go unrecognized.

Idiopathic DCM (IDCM) is characterized by an increased ventricular chamber size and reduced pumping of the heart (that is, reduced contractility; ref. 9 and Fig. 1) in the absence of coronary artery disease, valvular abnormalities or pericardial disease¹⁰. Backflow through the mitral valves and abnormal heart rhythms are common. Clinical features include common symptoms of CHF such as shortness of breath, easy fatigability, inability to tolerate physical exertion, fainting, light headedness, sweating at rest, and sudden death. Other signs are increased heart rate and an enlarged liver.

Most commonly, IDCM presents between 18 and 50 years of age, but children of all ages and the elderly can be affected. It occurs more frequently in men than in women (2.5:1), and in African-Americans than in Caucasians (2.5:1), but the causes of these differences are not understood¹¹. The clinical course of DCM, almost regardless of the underlying cause, may be progressive, with roughly 50% of individuals reported to die within 5 years of diagnosis without transplantation^{8,10}. The cause of death is divided evenly between sudden death and pump failure. Longer survival has been accomplished recently with improved medical therapies (such as angiotensin-converting enzyme

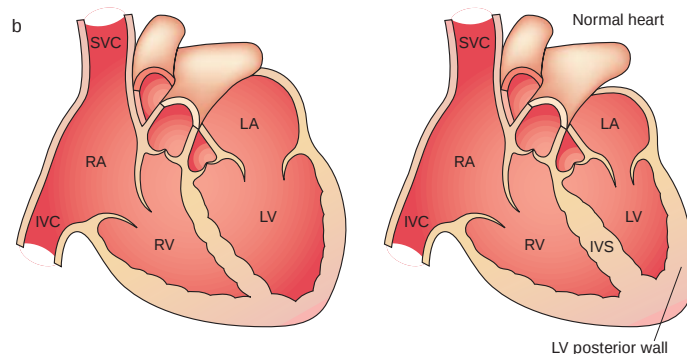
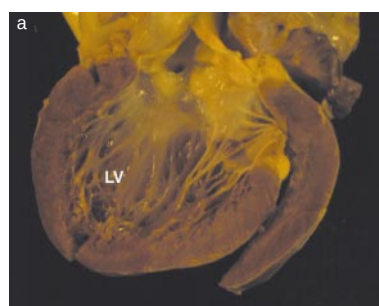


Figure 1 Dilated cardiomyopathy. **a**, Heart from an individual with dilated cardiomyopathy. Note the dilated left ventricle and thin ventricular walls. In life this ventricle pumped poorly. **b**, Illustration of dilated cardiomyopathy (left), showing a dilated left atrium and left ventricle, bulging interventricular

septum from left to right, and thin ventricular walls. For comparison, a normal heart is shown on the right. IVC, inferior vena cava; IVS, interventricular septum; LA, left atrium; LV, left ventricle; RA, right atrium; RV, right ventricle; SVC, superior vena cava.

inhibitors and β -blockers) and interventions (such as implantable defibrillators and ventricular assist devices).

Genetics of dilated cardiomyopathy

About 30–40% of individuals with DCM have a familial form of the disease (FDCM)¹². Autosomal dominant inheritance is the predominant pattern of transmission; X-linked, autosomal recessive and mitochondrial inheritance are less common^{9,12}. Mitochondrial inheritance is seen most often in childhood forms of FDCM, whereas X-linked and autosomal recessive forms seem to be evenly distributed between childhood and adult forms of disease.

Over the past decade, progress has been made in understanding the genetic aetiology of FDCM. Initial advances were made a decade ago through studies of families with X-linked forms, and in the past few years the autosomal dominant forms have become better understood. In the case of X-linked forms of DCM, two disorders have been well characterized: X-linked cardiomyopathy (XLCM), which presents in adolescence and young adults; and Barth syndrome, which is most frequently identified in infancy. The more common autosomal dominant form of FDCM also has two main forms: 'pure' DCM, and DCM associated with cardiac conduction system disease (CDDC).

Autosomal dominant dilated cardiomyopathy

Individuals who succumb to DCM with CDDC do so typically in their twenties, and what begins as mild conduction system disease can progress to complete heart block over decades. Dilated cardiomyopathy usually presents late in the course of disease but is out-of-proportion to the degree of conduction disorder. In many patients, mild conduction system disease occurs with subsequent development of severe DCM.

There is genetic heterogeneity of autosomal dominant DCM, with ten genetic loci mapped for pure DCM and five loci mapped for CDDC. In the case of pure DCM, the genetic loci identified so far include chromosomes 1q32, 2q31, 2q35, 4q12, 5q33, 9q13–22, 10q21–23, 14q11, 15q2 and 15q14 (refs 13–22). As listed in Table 1, seven of these genes are now known: *actin* (chromosome 15q14)²², *desmin* (chromosome 2q35)¹⁵, *δ -sarcoglycan* (chromosome 5q33)¹⁷, *β -sarcoglycan* (chromosome 4q12)¹⁶, *cardiac troponin T* (chromosome 1q32)²⁰, *β -myosin heavy chain* (chromosome 14q11)²⁰ and *α -tropomyosin* (chromosome 15q2)²¹.

Cardiac actin, a sarcomeric protein, is a member of the sarcomeric thin filament that interacts with tropomyosin and the troponin complex (Fig. 2). Actin is significant because it links the sarcomere to the sarcolemma by binding to the amino terminus of dystrophin (at the sarcolemma) and to β -myosin heavy chain (β -MyHC; in the sarcomere). Notably, DCM-related mutations of actin may affect directly its binding to dystrophin²²; by contrast, mutations in the sarcomeric end of actin result in hypertrophic cardiomyopathy (HCM)²³.

The DCM-causing mutations, first described by Olson *et al.*²², are thought to result in disease by causing force transmission abnormalities. The mutations in cardiac β -MyHC and cardiac troponin T are thought to cause reduced force generation by the sarcomere²⁰. In the case of β -MyHC, the location of the mutations seems to disrupt either stereospecific interactions between myosin and actin that are essential for initiating the power stroke of contraction, or a hinge region of myosin that transmits movement. By contrast, mutations in cardiac troponin T may reduce ionic interactions that form a tight binary complex of cardiac troponin T and cardiac troponin C, again reducing the power stroke of contraction²⁰. In individuals with

Figure 2 Proteins and pathways involved in the development of cardiomyopathies. Identified genes that result in dilated cardiomyopathy include the cytoskeletal protein-encoding genes β - and δ -sarcoglycan in the sarcolemma, *dystrophin* and the intermediate filament protein-encoding genes *desmin* and *lamin A/C*. Sarcomeric protein-encoding genes *actin*, *β -myosin heavy chain*, *α -tropomyosin* and *cardiac troponin T* cause either dilated cardiomyopathy or hypertrophic cardiomyopathy, whereas *cardiac troponin I*, *titin* and the myosin light chains cause hypertrophic cardiomyopathy. Mutations in *α -dystrobrevin* cause the intermediate phenotype, left ventricular noncompaction. MLP, muscle LIM protein; nNOS, neuronal nitric oxide synthase.

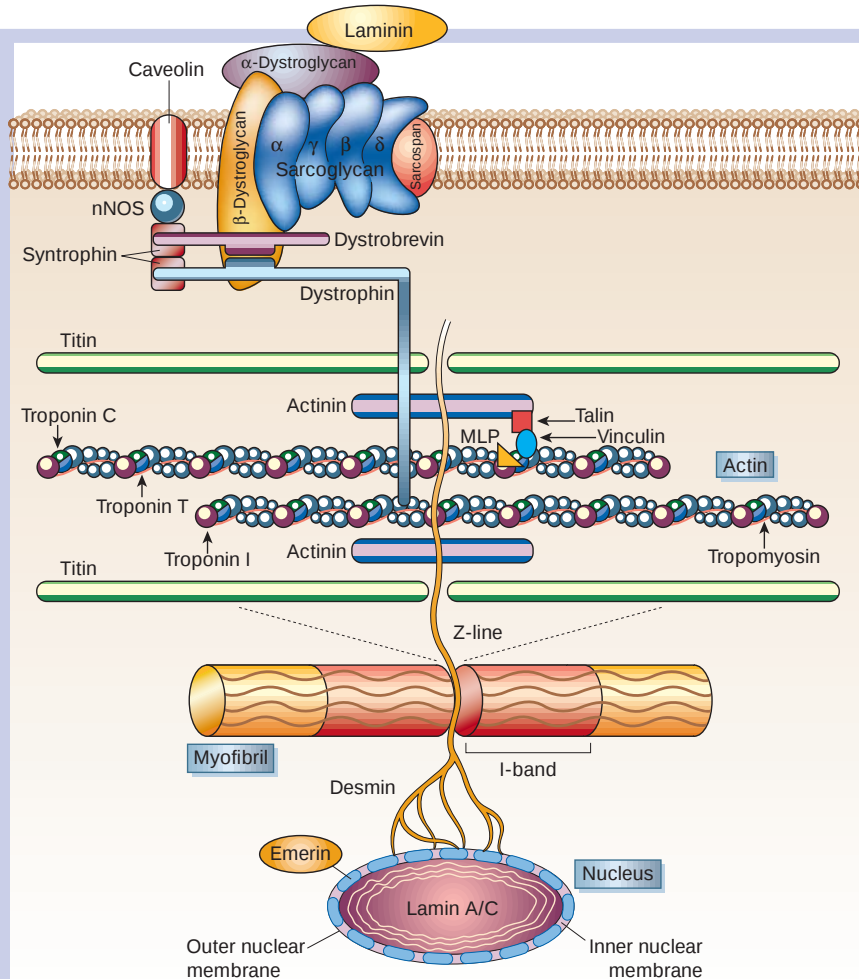


Table 1 Genetic causes of cardiomyopathies

Phenotype	Inheritance pattern	Chromosomal locus	Gene	Protein	Skeletal myopathy
Dilated cardiomyopathy	X-linked	Xp21	<i>dystrophin</i>	Dystrophin	Duchenne/Becker muscular dystrophy
	X-linked	Xq28	<i>G4.5</i>	Tafazzin	Barth syndrome
	Autosomal dominant	15q14	<i>actin</i>	Actin	Nemaline myopathy
		2q35	<i>desmin</i>	Desmin	Desmin myopathy
		5q33	<i>δ-sarcoglycan</i>	δ-Sarcoglycan	Limb girdle muscular dystrophy 2F
		1q32	<i>troponin T</i>	Troponin T	?
		14q11	<i>β-myosin heavy chain</i>	β-Myosin heavy chain	?
		15q2	<i>α-tropomyosin</i>	α-Tropomyosin	Nemaline myopathy
		MtDNA	<i>mitochondrial respiratory chain</i>	Mitochondrial respiratory chain	Mitochondrial myopathy
Dilated cardiomyopathy with conduction disease	Autosomal dominant	1q21	<i>lamin A/C</i>	Lamin A/C	Emery–Dreifuss muscular dystrophy
Hypertrophic cardiomyopathy	Autosomal dominant	14q11	<i>β-myosin heavy chain</i>	β-Myosin heavy chain	?
		1q32	<i>troponin T</i>	Troponin T	?
		12q23	<i>troponin I</i>	Troponin I	?
		15q2	<i>α-tropomyosin</i>	α-Tropomyosin	Nemaline myopathy
		11p11	<i>myosin-binding protein C</i>	Myosin-binding protein C	?
		3p21	<i>myosin essential light chain</i>	Myosin essential light chain	?
		3p21	<i>myosin regulatory light chain</i>	Myosin regulatory light chain	?
		2q31	<i>titin</i>	Titin	?
Hypertrophic cardiomyopathy with Wolff–Parkinson–White syndrome		7q3 MtDNA	<i>AMPK mitochondrial respiratory chain</i>	AMPK Mitochondrial respiratory chain	?
Left ventricular noncompaction	X-linked	Xq28	<i>G4.5</i>	Tafazzin	Barth syndrome
	Autosomal dominant	18q12	<i>α-dystrobrevin</i>	α-Dystrobrevin	Muscular dystrophy

DCM, mutations in α -tropomyosin²¹ are thought to affect the surface charge of the protein; these mutations may compromise the integrity of the thin filament, resulting in defects in force transmission²¹. In individuals with HCM, however, α -tropomyosin mutations show different characteristics²⁴.

Desmin is a cytoskeletal protein that forms intermediate filaments specific for muscle²⁵. It is found at the Z lines and intercalated disk of muscle, and its role in muscle function seems to involve attachment or stabilization of the sarcomere (Fig. 2). Mutations in this gene cause abnormalities of force and signal transmission that are similar to those thought to arise from mutations in actin¹⁵. Animal models show similar clinical features to those of humans with DCM²⁵.

δ-Sarcoglycan, a member of the sarcoglycan subcomplex of the dystrophin–glycoprotein complex (DGC; Fig. 2), is involved in stabilizing the myocyte sarcolemma and in signal transduction²⁶. Tsubata *et al.*¹⁷ have identified mutations in familial and sporadic cases of DCM that result in reduced protein within the myocardium. Animal models, such as the cardiomyopathic hamster and murine models, show a DCM phenotype^{27,28}. Another sarcoglycan complex member, β-sarcoglycan, can also cause DCM when mutated^{16,26}.

There is also genetic heterogeneity of CDDC. So far, CDDC genes have been mapped to chromosomes 1p1–1q1 (ref. 29), 2q14–21 (ref. 30), 3p22–25 (ref. 31) and 6q23 (ref. 32). The only gene that has been identified is *lamin A/C* on chromosome 1q21 (Table 1), which encodes a nuclear envelope intermediate filament protein (refs 33,34; and Fig. 2). The mechanisms responsible for the development of DCM and conduction system abnormalities are currently unknown.

X-linked dilated cardiomyopathy

First described in 1987 as a form of DCM occurring in males in their teens and early twenties with rapid progression from CHF to death or transplantation, XLCM is distinguished by elevated amounts of serum creatine kinase muscle isoforms³⁵ — a sign of underlying skeletal muscle disease. Female carriers tend to develop mild to moderate DCM in their fifties, and the disease is slowly progressive. Towbin *et al.*³⁶ identified the disease-causing gene, *dystrophin*, and showed that it is responsible for the clinical abnormalities owing to a severe reduction or absence of dystrophin protein in the heart of these individuals (Table 1). These findings were later confirmed, with most mutations clustering in the 5' portion of the gene and affecting the N-terminal actin-binding region of dystrophin³⁷.

Dystrophin is a cytoskeletal protein that provides structural support to myocytes by creating a lattice-like network to the sarcolemma³⁸. In addition, dystrophin has a principal role in linking the sarcomeric contractile apparatus to the sarcolemma and extracellular matrix^{39–41}. Furthermore, dystrophin is involved in cell signalling, particularly through its interactions with nitric oxide synthase⁴². The *dystrophin* gene is also responsible for Duchenne and Becker muscular dystrophy (DMD/BMD) when mutated³⁸. These skeletal myopathies present early in life (DMD is diagnosed before 12 years, whereas BMD is seen in teenage males older than 16 years), and most affected individuals develop DCM before 25 years of age. In most people with DMD/BMD, serum creatine kinase muscle isoforms are elevated to concentrations similar to those in XLCM; in addition, women carriers develop disease late in life, again similar to the X-linked condition. Furthermore, immunohistochemical analysis shows a reduced amount (or absence) of dystrophin, similar to that seen in the hearts of individuals with XLCM. Most individuals with DMD/BMD also develop DCM⁴³. It is an important disorder to understand because it is the most common skeletal myopathy, with a frequency of 1 in 3,500 newborn males. By contrast, the frequency of XLCM is not certain.

Murine models of dystrophin deficiency show abnormalities of muscle physiology that are caused by alterations in membrane structural support^{44,45}. In addition to the dysfunction of dystrophin, mutations in dystrophin secondarily affect proteins that interact with dystrophin. As mentioned above, the N terminus of dystrophin binds to the sarcomeric protein actin, a member of the thin filament of the contractile apparatus. The carboxy terminus of dystrophin interacts with β-dystroglycan, a dystrophin-associated membrane-bound protein that is involved in the function of the DGC, which includes α-dystroglycan, the sarcoglycan subcomplex (α-, β-, γ-, δ- and ε-sarcoglycan), syntrophins and dystrobrevins (refs 44,45; and Fig. 2). In turn, this complex interacts with α2-laminin and the extracellular matrix through α-dystroglycan. Similar to dystrophin, mutations in these proteins lead to muscular dystrophies with or without cardiomyopathy, supporting the idea that this group of proteins are important to the normal function of the myocytes of the heart and skeletal muscles^{41,43}. In both cases, mechanical stress⁴⁶ seems to be significant in the age-onset-dependent dysfunction of these muscles. The information gained from the studies on XLCM, DMD and BMD leads us to propose that DCM is a

primary disease of the cytoskeleton/sarcolemma, which leads to sarcomeric dysfunction^{47,48}.

Barth syndrome

Initially described as X-linked cardioskeletal myopathy with abnormal mitochondria and neutropenia (low white blood count), Barth syndrome typically presents in male infants as CHF associated with neutropenia (cyclic) and 3-methylglutaconic aciduria^{49,50}. Mitochondrial dysfunction is detected by electron microscopy and electron transport chain biochemical analysis. Echocardiograms show that these infants typically have left ventricular dysfunction with left ventricular dilation, endocardial fibroelastosis or a dilated hypertrophic left ventricle. In some cases they succumb to CHF/sudden death or sepsis caused by dysfunction of white blood cells. Most of these children survive infancy and do well clinically, although DCM usually persists.

The genetic basis of Barth syndrome was first described by Bione *et al.*⁵¹, who cloned the disease-causing gene *G4.5* (Table 1). This gene encodes the protein tafazzin, whose function is not known. But mutations in *G4.5* result in many clinical disorders⁵², including apparent classical DCM, hypertrophic DCM, endocardial fibroelas-

tosis and left ventricular noncompaction (LVNC)⁵³, with or without other features of Barth syndrome.

Animal models of dilated cardiomyopathy

So many of the genes identified for inherited DCM are also known to cause skeletal myopathies (Table 1), which suggests that skeletal and cardiac muscle share a 'final common pathway'. Further support for this concept comes from studies of animal models. Mutations in δ -sarcoglycan in hamsters result in cardiomyopathy²⁷, whereas mutations in any of the sarcoglycan subcomplex genes in mice cause skeletal and cardiac muscle disease^{28,44,45}. Mutations in other DGC genes, as well as in *dystrophin*, also consistently show abnormalities of skeletal and cardiac muscle function in murine models.

Arber *et al.*⁵⁴ have produced a mouse that is deficient in muscle LIM protein — a structural protein that links the actin cytoskeleton to the contractile apparatus. These mice develop severe DCM and CHF, and have disrupted cardiac myocyte cytoskeletal architecture. Badorff *et al.*⁵⁵ have shown that the DCM that develops after viral myocarditis may have a mechanism similar to the inherited forms. Using coxsackievirus B3 (CVB3) infection of mice, they showed that the CVB3-encoded protease 2A (enteroviral protease 2A) cleaves dystrophin, resulting in sarcolemmal disruption and possibly defects in force transmission, and DCM. A similar *dystrophin* mutation that affects individuals with XLCM shows a consistent mechanism of DCM development, namely, abnormalities of the cytoskeleton/sarcolemma.

Hypertrophic cardiomyopathy

Hypertrophic cardiomyopathy is a complex cardiac disease with unique pathophysiological characteristics and many morphological, functional and clinical features⁵⁶. In affected individuals the heart is thick; in particular, the interventricular septum and left ventricular posterior wall are thickened (usually asymmetric septal hypertrophy), and exaggerated pump function (that is, hypercontractile systolic function) is noted along with poor relaxation of the heart (diastolic dysfunction; Fig. 3). Obstruction of blood flow from the left ventricle to the aorta may also occur.

Although HCM has been considered by many to be a relatively uncommon cardiac disease, the prevalence of echocardiographically defined HCM in a large cohort of apparently healthy young adults selected from a community-based general population has been reported to be as high as 0.2% (ref. 56). Sporadic cases are common, but familial HCM (FHC) with autosomal dominant inheritance predominates⁵⁶. Hypertrophic cardiomyopathy seems to have a triphasic pattern of presentation⁵⁶. The most common form occurs in people aged between 10 and 25 years, but late-onset disease has also been reported and infants have been diagnosed with HCM in their first year of life. The underlying disease process in young children is thought to differ from that in older children and adults, with different apparent causes. Because HCM is a principal cause of sudden death in young, healthy individuals and athletes⁵⁶, it affects any population. In addition, a significant proportion of individuals develop CHF, caused either by diastolic dysfunction or by the development of left ventricular dilation and systolic dysfunction⁵⁶.

Genetics of familial hypertrophic cardiomyopathy

The first gene for FHC was mapped to chromosome 14q11.2–q12 using genome-wide linkage analysis in a large Canadian family⁵⁷. Soon afterwards, FHC locus heterogeneity was identified and several genes were mapped including chromosome 1q3, chromosome 15q2 and chromosome 11p11.2 (ref. 58). Five other loci were subsequently reported, located on chromosomes 7q3, 3p21.2–3p21.3, 12q23–q24.3, 15q14 and 2q31 (refs 23,58; and Table 1). Several other families are not linked with any known FHC loci, indicating the existence of additional FHC-causing genes.

The genes involved in FHC encode proteins that are part of the sarcomere—a complex structure with an exact stoichiometry and several sites of protein–protein interactions⁵⁹. The sarcomere proteins affected

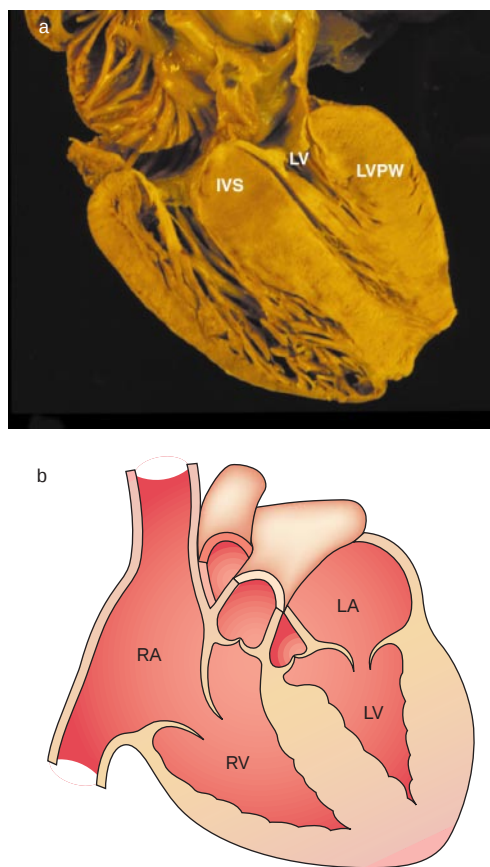
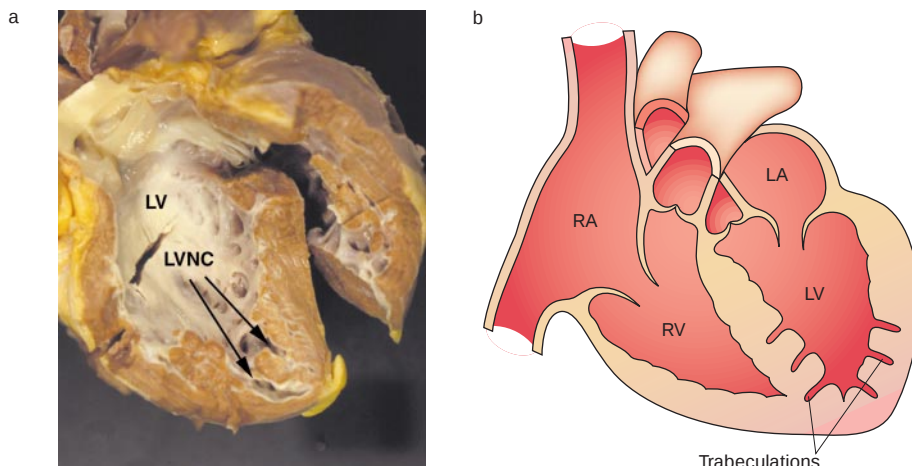


Figure 3 Hypertrophic cardiomyopathy. **a**, Heart from an individual with hypertrophic cardiomyopathy showing the severe thickening of the interventricular septum and posterior wall of the left ventricle and the tiny left ventricular chamber owing to encroachment of the thick walls. In life this ventricle was hypercontractile. IVS, interventricular septum; LVPW, left ventricular posterior wall. **b**, Illustration of hypertrophic cardiomyopathy. Note the severe thickening of the interventricular septum, which is thicker than the ventricular free wall (that is, asymmetric septal hypertrophy) and encroaches on the left ventricular chamber, which is small (compare with the normal heart shown in Fig. 1b, right).

Figure 4 Left ventricular noncompaction.

a, Heart from an individual with left ventricular noncompaction. Note the spongy appearance of the ventricular wall, caused by 'holes' in the myocardium, which represent deep trabeculations. The heart is thick and has a dilated chamber (that is, hypertrophic and dilated). In life this ventricle functioned poorly. LV, left ventricle; LVNC, trabeculations of left ventricular noncompaction. **b**, Illustration of left ventricular noncompaction. The left ventricular wall is thick and heavily trabeculated and the left ventricular chamber is dilated (compare with the normal heart shown in Fig. 1b, right).



by FHC include three myofilament proteins, β -MyHC, ventricular myosin essential light chain 1 (MLC-1s/v) and ventricular myosin regulatory light chain 2 (MLC-2s/v); four thin filament proteins, cardiac actin, cardiac troponin T, cardiac troponin I and α -tropomyosin (α -TM); one myosin-binding protein, the cardiac myosin-binding protein C (cMyBP-C); and titin (ref. 58 and Table 1). Each of these proteins is encoded by multigene families that show tissue-specific, developmental and physiologically regulated patterns of expression.

The gene responsible for cardiac hypertrophy and Wolff–Parkinson–White syndrome, previously linked to chromosome 7q3 (ref. 58), has been identified as that encoding the γ 2 subunit of AMP-activated protein kinase (AMPK), a gene that is important in energy use (refs 60,61; and Table 1). Affected individuals may develop disease in early life or as adults, and may succumb to CHF or sudden death. It is likely that these symptoms develop as a secondary result of sarcomeric dysfunction, and mutations in mitochondrial DNA⁹ result in dysfunction of respiratory chain enzymes of the mitochondrial energy-producing compartment and also lead to hypertrophic cardiomyopathy, especially in young children (Table 1). In some cases, hypertrophic dilated or frank dilated cardiomyopathy can develop. Hence, energy production and use, which is required for sarcomeric function and membrane integrity, are also important causes of these disorders.

Left ventricular noncompaction

Left ventricular noncompaction is characterized by deep trabeculations in the left ventricle endocardium (Fig. 4), in association with left ventricular hypertrophy, dilation or hypertrophy/dilation^{62,63}. In some individuals the left ventricle is hypercontractile, but systolic dysfunction predominates in others. Two forms of LVNC have been described, an isolated form and a form associated with congenital heart disease. In the latter case, this non-isolated LVNC may be found in combination with septal defects, pulmonic stenosis or hypoplastic left ventricle⁶³. The two forms of LVNC are found in all age groups, although an infantile presentation seems to be most common; survival seems to be good^{9,62,63}.

The inheritance patterns of LVNC include X-linked inheritance in males with isolated LVNC, and an autosomal dominant pattern in some isolated and non-isolated cases. In the X-linked form, mutations in the gene G4.5 (Table 1), which encodes tafazzin, were initially identified by Bleyl and co-workers^{62,63}. This is the gene identified by Bione and colleagues⁵³ as causative for Barth syndrome (see above); in fact, some individuals with LVNC also have clinical features of Barth syndrome⁶². Recently, we and our co-workers⁶³ identified a gene for non-isolated LVNC in a family with broad clinical features

ranging from apparent isolated LVNC, to LVNC with septal defects alone, to LVNC associated with hypoplastic left ventricle. This gene, α -dystrobrevin, encodes a dystrophin-associated protein (Fig. 2) that maps to chromosome 18q12 (refs 44,45; and Table 1). α -Dystrobrevin helps to maintain structural integrity of the muscle membrane, but also has signalling functions through the nitric oxide synthase pathway and is a substrate for tyrosine kinases^{44,45}. Deletion of this gene causes cardiomyopathy in mutant mice, supporting α -dystrobrevin as the cause of ventricular dysfunction^{44,45}. Again, skeletal myopathy is a feature to be noted. It is possible that the cardiac muscle abnormality is caused by disturbance of structural integrity and the congenital heart disease is due to perturbation of these signalling pathways in humans.

Animal models

Many animal models of HCM and LVNC have been generated⁶⁴. In the case of HCM, most of the known causative genes have been examined in mutant mice. The HCM phenotype first studied was that represented by the severe β -MyHC mutation Arg403→Gln, a mutation known to cause early death and severe hypertrophy in humans. Because MyHC mutations are responsible for 30–40% of the documented cases of HCM, MyHC represented a logical starting point for transgenic animal development. However, because the α -myosin heavy chain protein is predominant in mice (and not β -MyHC), the α -myosin heavy chain gene was mutated in the animal models and an Arg403→Gln mouse mutant was generated using a gene-targeting strategy⁶⁵. Homozygous animals were found to die shortly after birth, whereas sedentary heterozygotes survived, but showed abnormal cardiac structural and functional abnormalities typical of human HCM. Histopathological and physiological abnormalities included myocyte disarray, hypertrophy and fibrosis, which increased with age. Functional consequences included an acceleration in actin-activated cycling, increased rate of pressure rise and force development. Other MyHC mutations have also been generated, and typically the mutant mice show physiological abnormalities similar to those seen in humans carrying the corresponding mutations⁶⁴.

Murine models have also been developed for myosin light chain, myosin-binding protein C, α -tropomyosin, cardiac troponin T and troponin I, with variable similarities to the corresponding human mutations⁶⁴. Marian *et al.*⁶⁶ have developed a transgenic rabbit model caused by a common β -MyHC point mutation. As the predominant MyHC in rabbits is β -MyHC, mutant animals are more comparable to humans. In this model, wild-type and mutant human β -MyHC complementary DNAs were cloned 3' to a 7-kilobases murine

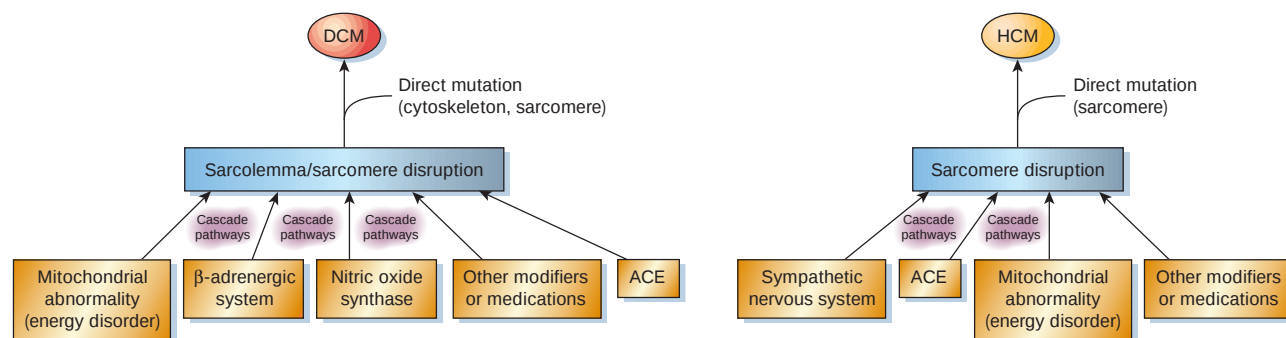


Figure 5 Final common pathways of dilated cardiomyopathy and hypertrophic cardiomyopathy. In dilated cardiomyopathy (DCM), the cytoskeleton linking sarcolemma and sarcomere seems to be the key pathway, with disruption of the linkage leading to dilation (through the sarcolemma) and poor pump function (through the sarcomere). By contrast, hypertrophic cardiomyopathy (HCM) is thought to occur primarily through sarcomere disruption in the face of intact cytoskeleton/sarcolemma. Cascade

pathways, such as mitochondrial (energy disruption) abnormalities, sympathetic tone or polymorphisms, signalling pathways, medications and other modifiers, can cause secondary disruption of these primary final common pathways that lead to the phenotype. Alternatively, direct mutations in the primary pathways can lead to the phenotype. ACE, angiotensin-converting enzyme.

β-MyHc promoter. Purified transgenes were injected into fertilized zygotes to generate two lines each of the wild-type and mutant transgenic rabbits. Animals carrying the mutant transgene showed substantial myocyte disarray, increased interstitial collagen expression in the myocardium, and hypertrophy of the interventricular septum and posterior wall of the left ventricle, mimicking the classical symptoms of human HCM. In addition, the incidence of premature death was increased. Therefore, this model is potentially more useful than mouse models for future studies to define mechanisms of human HCM and to develop new treatments.

Kittleson *et al.*⁶⁷ have reported a feline experiment of nature. A colony of Maine coon cats was found to have cardiomyopathy features characterized by moderate to severe papillary muscle and left ventricular concentric hypertrophy. In addition, the histopathology of the hearts of these animals has striking resemblance to human HCM, as does the natural history of the disease. Genetically, these cats have autosomal dominant inheritance with 100% penetrance. Crossbreeding the affected animals was lethal in these homozygotes, with stillborns having died *in utero*. Heterozygous animals developed HCM after six months of age, usually with an increasingly severe phenotype with age; classical features of HCM were seen typically during adolescence and young adulthood. Sudden death occurred in 20% of animals, and CHF was seen in 15% of cases. The gene responsible for this naturally occurring disease is not yet known.

Models of LVNC have been rarely described. So far, the only clearly defined model has been reported by Shou *et al.*⁶⁸, who used embryonic stem cells to generate mutant mice deficient in FK506-binding protein (FKBP12) — a protein that modulates the ryanodine receptor in its role as a calcium release channel. The FKBP12-deficient mice had normal skeletal muscle but a severely dilated form of LVNC. It is not clear how deficiency of the protein causes cardiac disease. Another model, the α-dystrobrevin-knockout mouse⁶⁹, develops hypertrophic dilated cardiomyopathy, although the authors did not describe features of LVNC. In this model, exon3-targeted deletion was used to generate the mutant mice; this is the exon that is mutated in individuals with LVNC. It is likely that mutation of this gene not only disrupts the DGC, but also impairs signalling through the nitric oxide pathway. Whether there is any relationship between α-dystrophin and FKBP12 is not known.

A common thread

Clearly, HCM of adults is a disease of the sarcomere²⁴. Similarly, FDCM is caused by mutations in genes encoding cytoskeletal

proteins^{47,48}. Hence, the final common pathways of these disorders include cytoskeletal proteins (DCM) and the sarcomere (HCM) (ref. 48; and Fig. 5). Similar pathways have been identified for ventricular arrhythmias (ion channels) and congenital heart disease (transcription factors)⁴⁸. In addition, it seems that whereas cascade pathways are involved directly in some cases (that is, mitochondrial abnormalities in HCM and DCM), secondary influences are likely to result in the various clinical symptoms seen in individuals with similar mutations.

In both DCM and HCM, mitochondrial and metabolic influences are also likely to be important. In addition, molecular interactions with such molecules as calcineurin, sex hormones, growth factors and other signalling pathways may be involved in the development of clinical signs, symptoms and age of presentation. In the future we expect these factors to be uncovered, allowing the development of new therapeutic strategies. Moreover, using these pathways for molecular screening is likely to uncover disease-causing genes in individuals with failing hearts.

The future for failure

With so much known about the genetic defects that underlie the cardiomyopathies, one obvious direction for the future treatment of this condition is gene therapy (see the review by Isner in this issue, pages 234–239). Although *dystrophin* is the best studied of the known DCM-causing genes, correcting the *dystrophin* defects associated with X-linked cardiomyopathy has not been attempted as yet. However, many groups have reported approaches for the phenotypic correction of the potentially lethal skeletal muscle disorder Duchenne muscular dystrophy. Adenovirus vectors encoding dystrophin have been used in *mdx* mice, resulting in efficient transduction of muscle fibres following intramuscular injection. Restoration of dystrophin-associated glycoprotein complex proteins to the sarcolemma and a decrease in centrally located nuclei, a characteristic of dystrophin-deficient skeletal muscle, was observed⁷⁰. As yet there have been no studies in the heart.

Although gene therapy may be many decades away, our increasing knowledge of the aetiological basis and mechanisms causing CHF is now starting to encourage different therapeutic approaches^{9,71}. In recent years it has become clear that therapies focused on improving cardiac function with inotropic agents improve the indices of contraction, as measured by echocardiography, but give no survival advantage. By contrast, β-blocker therapy^{9,71}, which reduces heart rate and is a negative inotropic agent, improves survival and leads to

reverse remodelling (that is, normalization of ventricular size and function). In addition, left ventricular assist devices can also lead to reverse remodelling⁷². These findings suggest that reducing myocardial mechanical stress can improve cardiac function, possibly through remodelling of the sarcomere–sarcolemma–extracellular matrix linkage. For example, dystrophin loss is reversed after 4–6 weeks of assist device therapy⁷³. In cases in which the dystrophin link is intact, administration of calcium channel blockers has been shown to be beneficial in mouse models^{74,75}. The benefit is thought to stem from protection from myocardial ischaemia but might result from calcium blockade or vasodilation⁷³. However, the identification of genetic aetiologies combined with tailored gene-specific therapeutics may offer significant improvements in disease prognosis and management in patients with congestive heart failure. □

1. O'Connell, J. B. & Bristow, M. R. Economic impact of heart failure in the United States: time for a different approach. *J. Heart Lung Transplant* **13**, S107–S112 (1994).
2. Richardson, P. et al. Report of the 1995 World Health Organization/International Society and Federation of Cardiology Task Force on the definition and classification of cardiomyopathies. *Circulation* **93**, 841–842 (1996).
3. Kushwaha, S. S., Fallon, J. T. & Fuster, V. Restrictive cardiomyopathy. *N. Engl. J. Med.* **336**, 267–276 (1997).
4. Corrado, D. et al. Spectrum of clinicopathologic manifestations of arrhythmogenic right ventricular cardiomyopathy/dysplasia: a multicenter study. *J. Am. Coll. Cardiol.* **30**, 1512–1520 (1997).
5. Thiene, G., Nava, A., Corrado, D., Rossi, L. & Pennelli, N. Right ventricular cardiomyopathy and sudden death in young people. *N. Engl. J. Med.* **318**, 129–133 (1988).
6. Furlanello, E. et al. Cardiac arrest and sudden death in competitive athletes with arrhythmogenic right ventricular dysplasia. *Pacing Clin. Electrophysiol.* **21**, 331–335 (1998).
7. Codd, M. B., Sugrue, D. D., Gersh, B. J. & Melton, L. J. Epidemiology of idiopathic dilated and hypertrophic cardiomyopathy. *Circulation* **80**, 564–572 (1989).
8. Abelman, W. H. & Lorrell, B. H. The challenge of cardiomyopathy. *J. Am. Coll. Cardiol.* **13**, 1219 (1989).
9. Towbin, J. A. in *The Molecular and Clinical Genetics of Cardiac Electrophysiological Disease* Ch. 13 (eds Berul, C. I. & Towbin, J. A.) 195–218 (Kluwer Academic, Norwell, MA, 2000).
10. Report of the 1995 World Health Organization/International Society and Federation of Cardiology Task Force on the Definition and Classification of Cardiomyopathies. *Circulation* **93**, 841–842 (1996).
11. Gillum, R. F. Idiopathic cardiomyopathy in the United States. *Am. Heart J.* **111**, 752–755 (1986).
12. Grunig, E. et al. Frequency and phenotypes of familial dilated cardiomyopathy. *J. Am. Coll. Cardiol.* **31**, 186–194 (1998).
13. Durand, J. B. et al. Localization of a gene responsible for familial dilated cardiomyopathy to chromosome 1q32. *Circulation* **92**, 3387–3389 (1995).
14. Siu, B. L. et al. Familial dilated cardiomyopathy locus maps to chromosome 2q31. *Circulation* **99**, 1022–1026 (1999).
15. Li, D. et al. Desmin mutations responsible for idiopathic dilated cardiomyopathy. *Circulation* **100**, 461–464 (1999).
16. Barresi, R. et al. Disruption of heart sarcoglycan complex and severe cardiomyopathy caused by β -sarcoglycan mutations. *J. Med. Genet.* **37**, 102–107 (2000).
17. Tsubota, S. et al. Mutations in the human δ -sarcoglycan gene in familial and sporadic dilated cardiomyopathy. *J. Clin. Invest.* **106**, 655–662 (2000).
18. Krajcinovic, M. et al. Linkage of familial dilated cardiomyopathy to chromosome 9. *Am. J. Hum. Genet.* **57**, 846–852 (1995).
19. Bowles, K. R. et al. Gene mapping of familial autosomal dominant dilated cardiomyopathy to chromosome 10q21–23. *J. Clin. Invest.* **98**, 1355–1360 (1996).
20. Kamisago, M. et al. Mutations in sarcomeric protein genes as a cause of dilated cardiomyopathy. *N. Engl. J. Med.* **343**, 1688–1696 (2000).
21. Olson, T. M., Kishimoto, N. Y., Whitby, F. G. & Michels, V. V. Mutations that alter the surface change of α -tropomyosin are associated with dilated cardiomyopathy. *J. Mol. Cell. Cardiol.* **33**, 723–732 (2001).
22. Olson, T. M., Michels, V. V., Thibodeau, S. N., Tai, Y. S. & Keating, M. T. Actin mutations in dilated cardiomyopathy, a heritable form of heart failure. *Science* **280**, 750–752 (1998).
23. Mogensen, J. et al. α -Cardiac actin is a novel disease gene in familial hypertrophic cardiomyopathy. *J. Clin. Invest.* **103**, R39–R43 (1999).
24. Thierfelder, L. et al. α -tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: a disease of the sarcomere. *Cell* **77**, 701–712 (1994).
25. Carlsson, L. & Thornell, L.-E. Desmin-related myopathies in mice and man. *Acta Physiol. Scand.* **171**, 341–348 (2001).
26. Hack, A. A., Groh, M. E. & McNally, E. M. Sarcoglycans in muscular dystrophy. *Microsc. Res. Tech.* **48**, 167–180 (2000).
27. Sakamoto, A. et al. Both hypertrophic and dilated cardiomyopathies are caused by mutation of the same gene, δ -sarcoglycan, in hamster: an animal model of disrupted dystrophin-associated glycoprotein complex. *Proc. Natl Acad. Sci. USA* **94**, 13873–13878 (1997).
28. Coral-Vazquez, R. et al. Disruption of the sarcoglycan-sarcospan complex in vascular smooth muscle: a novel mechanism for cardiomyopathy and muscular dystrophy. *Cell* **98**, 465–474 (1999).
29. Kass, S. et al. A gene defect that causes conduction system disease and dilated cardiomyopathy maps to chromosome 1p1–1q1. *Nature Genet.* **7**, 546–551 (1994).
30. Jung, M. et al. Investigation of a family with autosomal dominant dilated cardiomyopathy defines a novel locus on chromosome 2q14–q22. *Am. J. Hum. Genet.* **65**, 1068–1077 (1999).
31. Olson, T. M. & Keating, M. T. Mapping a cardiomyopathy locus to chromosome 3p22–p25. *J. Clin. Invest.* **97**, 528–532 (1996).
32. Messina, D. N., Speer, M. C., Pericak-Vance, M. A. & McNally, E. M. Linkage of familial dilated cardiomyopathy with conduction defect and muscular dystrophy to chromosome 6q23. *Am. J. Hum. Genet.* **61**, 909–917 (1997).
33. Fatkin, D. et al. Neonatal dilated cardiomyopathy in mice homozygous for the Arg403Gln mutant in the α -cardiac myosin heavy chain gene. *J. Clin. Invest.* **103**, 147–153 (1999).
34. Brodsky, G. L. et al. Lamin A/C gene mutation associated with dilated cardiomyopathy with variable skeletal muscle involvement. *Circulation* **101**, 473–476 (2000).
35. Berko, B. A. & Swift, M. X-linked dilated cardiomyopathy. *N. Engl. J. Med.* **316**, 1186–1191 (1987).
36. Towbin, J. A. et al. X-linked dilated cardiomyopathy (XLCMD): molecular genetic evidence of linkage to the Duchenne muscular dystrophy gene at the Xp21 locus. *Circulation* **87**, 1854–1865 (1993).
37. Muntoni, F. et al. Brief report: deletion of the dystrophin muscle-specific promoter region associated with X-linked dilated cardiomyopathy. *N. Engl. J. Med.* **329**, 921–925 (1993).
38. Hoffman, E. P., Brown, R. H. & Kunkel, L. M. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell* **51**, 919–928 (1987).
39. Meng, H., Leddy, J. J., Frank, J., Holland, P. & Tuana, B. S. The association of cardiac dystrophin with myofibrils/z-discs regions in cardiac muscle suggests a novel role in the contractile apparatus. *J. Biol. Chem.* **271**, 12364–12371 (1996).
40. Kaprielian, R. R., Stevenson, S., Rothery, S. M., Cullen, M. J. & Severs, N. J. Distinct patterns of dystrophin organization in myocyte sarcolemma and transverse tubules of normal and diseased human myocardium. *Circulation* **101**, 2586–2594 (2000).
41. Campbell, K. P. Three muscular dystrophies: loss of cytoskeleton–extracellular matrix linkage. *Cell* **80**, 675–679 (1995).
42. Chang, W. J. et al. Neuronal nitric oxide synthase and dystrophin-deficient muscular dystrophy. *Proc. Natl Acad. Sci. USA* **93**, 9142–9147 (1996).
43. Cox, G. F. & Kunkel, L. M. Dystrophies and heart disease. *Curr. Opin. Cardiol.* **12**, 329–343 (1997).
44. Allamand, V. & Campbell, K. P. Animal models for muscular dystrophy: valuable tools for the development of therapies. *Hum. Mol. Genet.* **9**, 2459–2467 (2000).
45. Heydemann, A., Wheeler, M. T. & McNally, E. M. Cardiomyopathy in animal models of muscular dystrophy. *Curr. Opin. Cardiol.* **16**, 211–217 (2001).
46. Petrof, B. J., Shrager, J. B., Stedman, H. H., Kelly, A. M. & Sweeney, H. L. Dystrophin protects the sarcolemma from stresses developed during muscle contraction. *Proc. Natl Acad. Sci. USA* **90**, 3710–3714 (1993).
47. Towbin, J. A. The role of cytoskeletal proteins in cardiomyopathies. *Curr. Opin. Cell Biol.* **10**, 131–139 (1998).
48. Bowles, N. E., Bowles, K. R. & Towbin, J. A. The “Final Common Pathway” hypothesis and inherited cardiovascular disease: the role of cytoskeletal proteins in dilated cardiomyopathy. *Herz* **25**, 168–175 (2000).
49. Barth, P. G. et al. An X-linked mitochondrial disease affecting cardiac muscle, skeletal muscle and neutrophil leukocytes. *J. Neurol. Sci.* **62**, 327–355 (1983).
50. Kelley, R. I. et al. X-linked dilated cardiomyopathy with neutropenia, growth retardation, and 3-methylglutaconic aciduria. *J. Pediatr.* **119**, 738–747 (1991).
51. Bione, S. et al. A novel X-linked gene, G4.5, is responsible for Barth syndrome. *Nature Genet.* **12**, 385–389 (1996).
52. D'Adamo, P. et al. The X-linked gene G4.5 is responsible for different infantile dilated cardiomyopathies. *Am. J. Hum. Genet.* **61**, 862–867 (1997).
53. Bleyl, S. B. et al. Neonatal, lethal noncompaction of the left ventricular myocardium is allelic with Barth syndrome. *Am. J. Hum. Genet.* **61**, 868–872 (1997).
54. Arber, S. et al. MLP-deficient mice exhibit a disruption of cardiac cytoarchitectural organization, dilated cardiomyopathy, and heart failure. *Cell* **88**, 393–403 (1997).
55. Badorf, C. et al. Enteroviral protease 2A cleaves dystrophin's evidence of cytoskeletal disruption in an acquired cardiomyopathy. *Nature Med.* **5**, 320–326 (1999).
56. Maron, B. J. Hypertrophic cardiomyopathy. *Lancet* **350**, 127–133 (1997).
57. Jarcho, J. A. et al. Mapping a gene for familial hypertrophic cardiomyopathy to chromosome 14q1. *N. Engl. J. Med.* **321**, 1372–1378 (1989).
58. Towbin, J. A. Molecular genetics of hypertrophic cardiomyopathy. *Curr. Cardiol. Rep.* **2**, 134–140 (2000).
59. Thierfelder, L. et al. A familial hypertrophic cardiomyopathy locus maps to chromosome 15q2. *Proc. Natl Acad. Sci. USA* **90**, 6270–6274 (1993).
60. Mann, C. J. et al. Antisense-induced exon skipping and synthesis of dystrophin in the mdx mouse. *Proc. Natl Acad. Sci. USA* **98**, 42–47 (2001).
61. Ahmad, A., Brinson, M., Hodges, B. L., Chamberlain, J. S. & Amalfitano, A. Mdx mice inducibly expressing dystrophin provide insights into the potential of gene therapy for Duchenne muscular dystrophy. *Hum. Mol. Genet.* **9**, 2507–2515 (2000).
62. Nowak, K. J. et al. Mutations in the skeletal muscle α -actin gene in patients with actin myopathy and nemaline myopathy. *Nature Genet.* **23**, 208–212 (1999).
63. Ichida, F. et al. Novel gene mutations in patients with left ventricular noncompaction or Barth syndrome. *Circulation* **103**, 1256–1263 (2001).
64. Dalloz, F., Osinska, H. & Robbins, J. Manipulating the contractile apparatus: genetically defined animal models of cardiovascular disease. *J. Mol. Cell. Cardiol.* **33**, 9–25 (2001).
65. Fatkin, D. et al. Neonatal dilated cardiomyopathy in mice homozygous for the Arg403Gln mutation in the α -cardiac myosin heavy chain gene. *J. Clin. Invest.* **103**, 147–153 (1999).
66. Marian, A. J. et al. A transgenic rabbit model for human hypertrophic cardiomyopathy. *J. Clin. Invest.* **104**, 1683–1692 (1999).
67. Kittleson, M. D. et al. Familial hypertrophic cardiomyopathy in Maine Coon cats. An animal model of human disease. *Circulation* **99**, 3172–3180 (1999).
68. Shou, W. et al. Cardiac defects and altered ryanodine receptor function in mice lacking FKBP12. *Nature* **391**, 489–492 (1998).
69. Grady, R. M. et al. Role for α -dystrobrevin in the pathogenesis of dystrophin-dependent muscular dystrophies. *Nature Cell Biol.* **1**, 215–220 (1999).
70. Clemens, P. R. et al. In vivo muscle gene transfer of full-length dystrophin with an adenoviral vector that lacks all viral genes. *Gene Ther.* **3**, 965–972 (1996).
71. Braunwald, E. & Bristow, M. R. Congestive heart failure: fifty years of progress. *Circulation* **102**, IV-14–IV-23 (2000).
72. Deng, M. C. et al. Mechanical circulatory support for advanced heart failure. Effect of patient selection or outcome. *Circulation* **103**, 231–237 (2001).
73. Stetson, S. J. et al. Evidence for reversible dystrophin abnormalities in patients with non-familial dilated cardiomyopathies: observations in patients treated with long-term mechanical support. *Circulation* **102**, II-132 (2000).
74. Cohn, R. D. et al. Prevention of cardiomyopathy in mice lacking smooth muscle sarcoglycan-sarcospan complex. *J. Clin. Invest.* **107**, R1–R7 (2001).
75. Towbin, J. A. & Bowles, N. E. Sarcoglycan, the heart, and skeletal muscles: new treatment, old drug? *J. Clin. Invest.* **107**, 154–154 (2001).

Myocardial gene therapy

Jeffrey M. Isner*

Department of Medicine, Divisions of Cardiology, Vascular Medicine, and Cardiovascular Research, St Elizabeth's Medical Center, Tufts School of Medicine, Boston, Massachusetts 02135, USA

Gene therapy is proving likely to be a viable alternative to conventional therapies in coronary artery disease and heart failure. Phase 1 clinical trials indicate high levels of safety and clinical benefits with gene therapy using angiogenic growth factors in myocardial ischaemia. Although gene therapy for heart failure is still at the pre-clinical stage, experimental data indicate that therapeutic angiogenesis using short-term gene expression may elicit functional improvement in affected individuals.

After a decade of pre-clinical and early phase 1 clinical investigations, gene therapy has emerged as a genuine therapeutic option with the potential to alter the manner in which cardiologists manage the two most common cardiac disorders afflicting adults — coronary artery disease and congestive heart failure. Here I provide insights into the application of gene transfer to ischaemic heart disease and heart failure, based on currently available vectors, delivery instruments, transgenes, and what is currently known regarding the safety and efficacy of gene therapy in human clinical trials.

Myocardial ischaemia

The idea that angiogenic growth factors might promote revascularization of ischaemic tissues — a strategy termed 'therapeutic angiogenesis'¹ — was first investigated in individuals with peripheral artery disease, specifically critical limb ischaemia². Gene transfer of plasmid DNA encoding vascular endothelial growth factor (VEGF) brought about clinical benefits, such as the abolition of rest pain, limb salvage and the healing of ischaemic ulcers. These benefits were associated with angiographic evidence of new collateral vessels and also improved leg blood flow as monitored by magnetic resonance angiography^{3,4}. The concept of therapeutic angiogenesis in human subjects thus proved, phase 1 clinical trials involving different gene transfer strategies were undertaken to test this scheme in people affected with myocardial ischaemia⁵.

Choice of vector and gene

In contrast to non-cardiovascular applications of gene therapy, most of which use viral vectors including adenovirus, adeno-associated virus, retrovirus and lentivirus, cardiovascular trials are remarkable for their use of non-viral vectors. Forty-seven per cent of these trials used either naked plasmid DNA (39%) or liposome carriers (8%). The disproportionate use of non-viral vectors runs counter to the notion that the use of naked DNA is too inefficient to be an effective strategy for gene transfer. This notion prevailed for most of the first decade of gene therapy, and attention was instead focused on the use of viral vectors, which were intended to increase the magnitude and/or duration of gene expression.

Genes encoding proteins that must remain intracellular to achieve a biological effect have to be delivered to a

relatively large target population of cells to correct the underlying pathogenetic defect. In contrast, genes encoding proteins that are naturally secreted can achieve favourable effects when limited numbers of cells are transfected, provided that the transfected cells secrete substantial amounts of the gene product⁶. The paracrine effect of the secreted gene product can then modulate the bioactivity of several target cells.

Analysis of the transgenes that have been selected for cardiovascular trials indicates a preference for vectors encoding angiogenic growth factors (for example, VEGF^{5,7} and fibroblast growth factor-4 (FGF-4; ref. 8)) that have a native or ligated signal sequence permitting active secretion.

Improving the odds

Some technical factors related to the mode of gene delivery also contribute to successful gene expression. First, as initially shown by Wolff *et al.*⁹, muscle, and in particular ischaemic muscle¹⁰, is highly permissive to the transfer of plasmid DNA. This might be due to its multinucleate histology, which provides several targets per cell for transgene nuclear transfer. Second, gene expression can be augmented as a direct function of the volume of injected excipient, independent of the dose of DNA¹¹. Transcutaneous monitoring of individuals undergoing gene transfer has shown that an aliquot of 2.5 ml is typically distributed over a distance of 5.8 cm (ref. 12); and *in vivo* studies in laboratory swine have reported redistribution of the plasmid throughout the myocardium — away from the site of injection — after intramuscular gene transfer¹³. Third, intraoperative, epicardial gene transfer has been facilitated by the use of trans-oesophageal echocardiographic guidance, which enables cardiologists to explore the heart from the oesophagus and stomach¹⁴. This ensures that the intramyocardial injection is performed properly into the myocardium and not compromised by inadvertent injection into the ventricular cavity, where the transgene would be diluted and degraded in circulating blood.

Last, direct intramyocardial gene transfer may be optimized further by modifications in the backbone vector¹⁵ and/or the mode of physical delivery to achieve meaningful clinical results while minimizing safety concerns. For example, Schratzberger *et al.*¹⁶ have shown that ultrasound radiation (1 MHz, 6% duty cycle) applied immediately after intramuscular injection permeabilizes cell membranes to enhance uptake of plasmid DNA and may augment gene expression by more than 20-fold. Taniyama *et al.*¹⁷ have shown that adjunctive use of ultrasound exposure increases the efficiency of naked DNA

*We regret the sad news that Dr Jeffrey Isner died suddenly during the latter stages of preparing his Insight Review Article. *Nature's* editors and colleagues of Dr Isner agreed that his review should be published, and permission to proceed has kindly been given by Dr Isner's next-of-kin. *Nature* carries sole responsibility for editorial changes made to the article during the production process.

gene transfer to the point that meaningful biological responses can be achieved with even non-secreted gene products such as p53.

Transient expression

Long-term gene expression, which can be achieved with non-viral plasmid DNA or adenoviral vectors, may be a prerequisite for some types of inherited genetic defects, but does not seem to be a requirement for gene transfer strategies designed to promote neovascularization¹⁸.

The safety profile of short-term or 'transient expression' has been verified in initial trials of gene transfer for myocardial ischaemia. Out of 85 affected individuals participating in four protocols using naked DNA VEGF gene transfer, there have been three deaths. And five deaths have occurred out of 97 individuals that have been followed for 1–3 years after myocardial gene transfer of adenovirus encoding either FGF-4 or the 121 isoform of the VEGF-1 gene (VEGF₁₂₁; K. Hammond, unpublished data). These outcomes compare favourably with the mortality rate in a similar group of almost 1,000 individuals receiving laser myocardial revascularization or contemporary medical therapy¹⁹.

Although clinical investigations of gene transfer for myocardial ischaemia have so far consisted exclusively of phase 1 trials designed mainly to address safety concerns, intriguing hints of bioactivity have been observed. For example, out of 30 individuals receiving plasmid DNA encoding the 165-amino-acid isoform of VEGF (phVEGF₁₆₅),

29 experienced reduced angina and sublingual consumption of nitroglycerin²⁰. Moreover, angina was eliminated altogether in 12/20 individuals who were followed for more than 6 months, and in 7/10 individuals followed for over 12 months. Exercise tolerance, as measured by the Bruce protocol, increased significantly (240 s before gene transfer, compared with 410 s up to 180 d after gene transfer, $P < 0.05$). In addition, the left ventricular ejection fraction was either unchanged ($n = 16$) or improved ($n = 14$; mean increase 5%) after phVEGF₁₆₅ gene transfer²⁰.

Our own experience with intramyocardial naked DNA gene transfer has been reproduced in a multicentre, consecutive series, dose-escalating study of transepical VEGF-2 gene transfer in 30 individuals suffering with class 3 or 4 angina. Once again, the frequency of angina attacks and consumption of nitroglycerin were both reduced, and exercise time was increased by over 2 min (J. M. I., unpublished data). Of the 30 individuals, 29 are alive after a minimum of 12 months follow-up, and none has undergone cardiac transplantation. A single post-operative death occurred 20 h after gene transfer.

In contrast to the use of plasmid DNA, three other clinical trials have investigated the use of myocardial gene transfer using adenoviral vectors. First, Rosengart *et al.*⁷ delivered 4×10^8 to 4×10^{10} particle units of a first generation adenoviral vector encoding the VEGF₁₂₁ gene to 15 individuals undergoing bypass graft surgery, and as sole therapy to six individuals by means of mini-thoracotomy. There were no adverse events related to vector administration, and symptoms improved in both groups. The same investigators have since injected ten individuals with adenoviral VEGF₁₂₁ transepically using video-assisted thoracotomy or thoracoscopy, with one late death²¹; the symptomatic outcomes are yet to be reported.

Second, the preliminary results of a multicentre, double-blind, randomized trial of direct intracoronary administration of adenovirus encoding FGF-4 (AdFGF-4) have been reported recently⁸. Incremental doses of 3×10^8 to 1×10^{11} AdFGF-4 particles were administered to 67 individuals (at a 3:1 ratio of active to placebo) suffering with mild to moderate angina. Although the overall improvement in exercise treadmill time (ETT) was similar for individuals receiving AdFGF-4 or placebo, post-hoc analyses revealed potentially important differences related to their baseline neutralizing antibody titre to Ad5. Among 46 individuals with titres of less than 1:100, 44% had increased their ETT time by more than 30% at the week 12 follow-up; in contrast, among 14 individuals with titres greater than 1:100, 7% had increased their 12-week ETT by more than 30%.

Third, an adenoviral vector is being investigated to achieve gene transfer of a constitutively active form of hypoxia inducible factor 1 α (HIF-1 α)²². In this initial phase 1 clinical trial, AdHIF1 α /VP16 is injected directly into a region of left ventricular myocardium that cannot be successfully revascularized by conventional means in individuals undergoing coronary artery bypass surgery. Clinical outcomes for this initial trial have not yet been disclosed. *In vitro* analyses have shown that HIF-1 α promotes the upregulation of endogenous VEGF gene expression in HeLa and C6 cells, as well as the upregulation of both VEGF and erythropoietin in Hep3B cells independent of hypoxia. HIF-1 α also augments collateral vessel development in both rabbits with hindlimb ischaemia²² and a swine model of ameroid constrictor-induced myocardial ischaemia²³.

The first clinical study of percutaneous, catheter-based (Fig. 1) myocardial gene transfer has been completed²⁴. This single-blind pilot study, designed to establish the safety and feasibility of non-surgical needle injection, involved a total of six individuals randomized to either gene transfer or a mock procedure; each of the three individuals assigned to the mock group ultimately required crossover to gene transfer because of persistent symptoms. The number of angina attacks fell similarly in the active and control groups over the first 30 days after gene transfer, suggesting an initial placebo effect. But

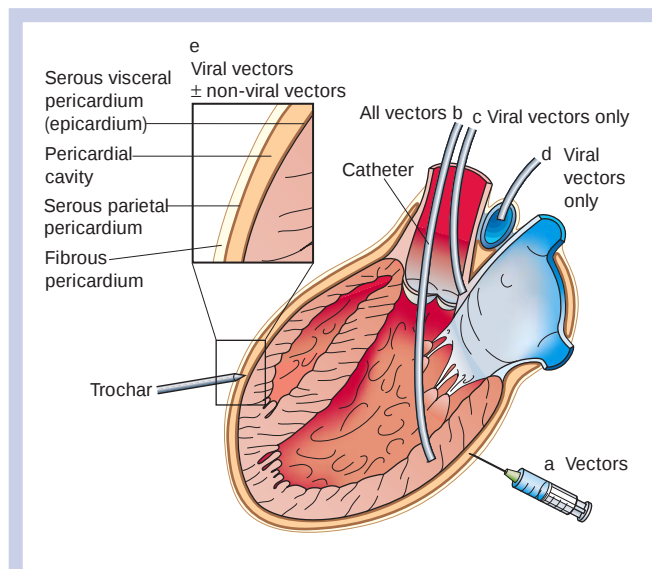


Figure 1 Delivery options for implementing myocardial gene transfer. **a**, Epicardial. After intraoperative intramyocardial injection, the transgene is distributed to local cardiomyocytes and/or myocardial interstitium, with a gradient of redistribution that becomes more diminutive as a function of distance from site of injection. **b**, Endocardial. After catheter-based intramyocardial injection, the transgene is distributed in a similar way to that described in **a**. **c**, Intracoronary. Catheter injection directly into coronary arteries delivers the transgene by means of the coronary arterial circulation to a broad sample of cardiomyocytes and/or myocardial interstitium in the area subserved by the injected, source artery; note that this approach requires a patent artery or bypass conduit. **d**, Retroperfusion. Cannulation of a coronary sinus (or subselective vein) with concurrent occlusion of outflow from the coronary sinus distributes the transgene retrograde from coronary venous circulation through the capillary bed to cardiomyocytes and interstitium in potentially all regions of left ventricular myocardium. **e**, Intrapericardial. Transthoracic access to pericardium using a specialized catheter permits delivery into the pericardial space with secondary distribution directly to epicardial cardiomyocytes, and indirectly through the circulation shared by the peri- and epicardium. As indicated, options **b** and **c** are not appropriate for non-viral gene transfer owing to intracirculatory degradation of plasmid DNA unprotected by the viral vector.

longer follow-up disclosed that angina continued to decrease at 90 days only in the active group; in contrast, individuals in the mock group regressed towards their initial symptomatic status, suggesting that this late reduction in angina was not a placebo effect. The symptomatic improvement was again accompanied by non-invasive evidence of improved myocardial perfusion, and electromechanical evidence of improved myocardial function. This study has served as the basis for a multicentre, randomized, double blind, placebo-controlled trial of catheter-based VEGF-2 gene transfer that has so far enrolled 19 individuals.

Heart failure: viral vectors

The elegant dissection of the molecular pathways that contribute to heart failure has provided the theoretical underpinnings for a strategy of “molecular ventricular assistance”²⁵ to address human heart failure (Fig. 2). Initial experimental applications have evaluated specific molecular interventions in generic models of heart failure (for example, coronary ligation, microsphere embolization and ventricular pacing). Successful application of these ‘one size fits all’ treatments is encouraging with respect to people with heart failure — most of whom are currently ‘lumped’ together as clinical heart failure without further molecular distinction. But the ultimate evolution of these experimental studies, in combination with genomic and proteomic studies, may be the potential for gene transfer strategies that are customized to target the individual molecular defects²⁶ responsible for heart failure.

Although studies have yet to be initiated in human subjects, three types of gene transfer studies designed to enhance myocardial contractility directly have been investigated in pre-clinical studies. These include modulating calcium homeostasis, manipulating β -adrenergic receptor (β -AR) signalling, and augmenting cardiomyocyte resistance to apoptosis.

Calcium homeostasis

On the basis of earlier evidence showing that reduced expression or activity of the sarcoplasmic reticulum Ca^{2+} -ATPase (SERCA) leads to increased concentrations of cytosolic Ca^{2+} and less available Ca^{2+} for release from the sarcoplasmic reticulum²⁷, Miyamoto *et al.*²⁸ showed that adenoviral overexpression of SERCA2a restored both systolic and diastolic function to normal levels in a rat model of pressure-overload hypertrophy in transition to heart failure. Specifically, the left ventricle was reduced in size, and the slope of the end systolic pressure/dimension relationship (a load-independent parameter of contractility) was restored to control levels. Moreover, SERCA2a expression decreased diastolic Ca^{2+} — a result that might be expected to prevent activation of signalling molecules, including calcineurin and stress-activated protein kinases capable of inducing myocyte hypertrophy and cell death²⁹.

Similar findings in studies of transgenic mice³⁰ further support the concept that overexpression of SERCA2a can successfully rescue disturbed Ca^{2+} cycling and myocardial function in the failing heart, although SERCA overexpression in gene transfer or transgenic studies has not exceeded 1.5- to 2-fold³¹. More recently, overexpression of an antisense phospholamban construct or a dominant-negative mutant of phospholamban — an endogenous muscle-specific inhibitor of the sarcoplasmic reticulum calcium pump — has been shown to enhance SERCA2 activity³².

β -Adrenergic receptors

The well-documented abnormalities of β -AR signalling³³ comprise a robust molecular target for gene transfer in people with heart failure. These abnormalities include downregulation of β -ARs, uncoupling of second-messenger systems, and upregulation of β -AR kinase. Maurice *et al.*³⁴ have shown the feasibility of performing intracoronary injection of 5×10^{11} viral particles encoding β -AR in rabbits with normal hearts. Gene expression was increased for up to 3 weeks; basal and isoproterenol-stimulated contractile function were both

increased significantly, and the latter persisted throughout the 3-week duration of enhanced expression.

Weig *et al.*³⁵ have used a new strategy to circumvent desensitized β -ARs by performing adenoviral (10^{10} p.f.u.) gene transfer of V2 vasopressin receptors, which have been shown previously to promote activation of adenylyl cyclase in kidney. Successful heterologous expression of V2 receptors in rat and rabbit myocardium produced increased fractional shortening in rats and rabbits, and increased dP/dt_{max} in rats.

β -ARK1, a member of the G-protein-coupled receptor kinase family (see review in this issue by Rockman, Koch and Lefkowitz, pages 206–212), phosphorylates agonist-occupied receptors such as β -ARs, which leads to maladaptive receptor desensitization³⁶. Shah *et al.*²⁵ created a rabbit model of heart failure by ligating the obtuse marginal branch of the left circumflex coronary artery. Three weeks later, they injected 5×10^{11} viral particles of adenovirus encoding a peptide inhibitor of β -ARK1 through a catheter into the parent circumflex artery. Significant improvement in measures of left ventricular systolic function were observed, as well as a trend towards increased left ventricular dP/dt_{max} . As expected, the peptide inhibitor did not alter concentrations of β -ARK1, but improved β -AR signalling.

A potential liability of this approach is sustained adrenergic stimulation, which may be arrhythmogenic and associated with increased mortality. Lai *et al.*³⁷ attempted to obviate this concern by the intracoronary adenoviral gene transfer of 1.4×10^{12} adenoviral particles encoding adenylyl cyclase VI (AC_{VI}). The concentration of AC_{VI} protein increased, and a twofold increase in cyclic-AMP-generating capacity persisted for up to 18 weeks; these findings were associated with a statistically significant increase in left ventricular dP/dt_{max} and cardiac output during stimulation with isoproterenol. Importantly, however, both resting haemodynamics and resting adenylyl cyclase activity were reportedly unchanged; thus, the method seems to address successfully the potential jeopardy associated with persistent catecholamine stimulation of the myocardium.

Resisting apoptosis

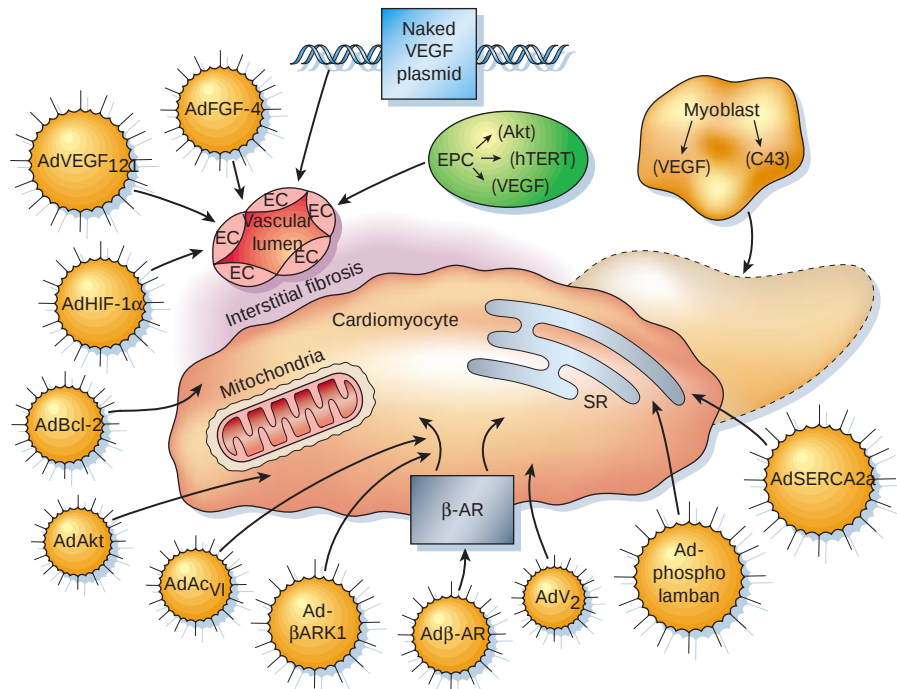
Several groups have demonstrated cardiomyocyte apoptosis in specimens from humans with heart failure, leading to the proposal that gene transfer techniques that promote myocyte survival may offer therapeutic benefit³⁸. Conceptual support for this idea is derived from mice that have undergone ventricular-restricted knockout of gp130; these mice develop rapid-onset, dilated cardiomyopathy associated with massive cardiomyocyte apoptosis³⁹. Although *in vivo* studies that test this idea have not been reported, *in vitro* experiments have reported favourable effects on cardiomyocyte survival after adenoviral gene transfer of Bcl-2 (ref. 40) and Akt^{41,42} to block apoptosis induced by p53 and hypoxia, respectively.

Adenoviral vectors

Adenoviral vectors have been used to good advantage by several laboratories to prove the concept of various gene transfer strategies in animal models, but it is unlikely that this approach will have use in people with heart failure, because of the limited duration of gene expression associated with this vector. This problem is compounded by the uncertainty of repeat administration of adenoviral vectors. Cardiovascular trials of gene therapy using viral vectors (all adenovirus) have established that neutralizing antibodies directed against first generation adenoviral vectors are increased in most⁸ or all⁷ individuals, and may thus diminish the impact of subsequent treatments.

Although some pre-clinical studies have been interpreted to suggest that inflammation provoked by the use of viral vectors increases the risk of gene transfer, other pre-clinical studies have failed to show evidence of an inflammatory reaction^{43,44}. In this regard, Rosengart *et al.*⁷ reported no increase in serum CPK or other findings indicative of myocarditis in individuals receiving adenoviral VEGF₁₂₁ gene transfer as sole therapy, and the same was true for adenoviral administered FGF-4. But *in vitro* studies by O'Donnell *et al.*⁴⁵ have underscored the

Figure 2 Strategies of gene transfer for treatment of heart failure. Cellular targets for myocardial gene transfer include cardiomyocytes and endothelial cells (EC). Transduction of cardiomyocytes is facilitated by use of viral vectors. The transgenes indicated here have been used successfully in pre-clinical *in vivo* studies to modulate calcium homeostasis at the level of the sarcoplasmic reticulum (SR), to manipulate β -adrenergic receptor (β -AR) signalling, and to augment resistance to apoptosis. In addition, pre-clinical studies have also established that it is possible to enhance the blood flow to viable cardiomyocytes encased by excessive interstitial fibrosis by using both non-viral and viral vectors to deliver transgenes encoding angiogenic growth factors to promote microcirculatory neovascularization. Finally, indirect gene transfer may target the cellular substrate by *ex vivo* transfection of either endothelial progenitor cells (EPC) or skeletal myoblasts before *in vivo* cellular transplantation.



crucial issue of dose range, showing that — at least for rat cardiomyocytes — viral titres greater than 5–6 p.f.u. per cell result in apoptotic cell death.

The clinical issues associated with adenoviral vectors may be addressed more successfully with alternative viral vectors, particularly adeno-associated virus (AAV) or lentiviral vectors. In animal models, AAV has been shown to yield widespread gene expression of reporter genes⁴⁶, and there is evidence of enhanced myocardial revascularization after AAV-VEGF gene transfer⁴⁷. Although AAV vectors are more difficult to generate in high titres, the promise of both AAV and lentivirus is tied to preliminary indications of reduced immunogenicity and far more protracted gene expression, probably owing to the ability of these vectors to integrate into the host genome with a presumably low risk of insertional mutagenesis. The corollary complication of protracted, unrestricted gene expression is, however, the potential for deleterious side-effects, as illustrated by the example of elevated resting concentrations of adenylyl cyclase cited above. It is this particular concern that has caused others to suggest the possible need for vectors that can be regulated⁴⁸.

Heart failure: plasmid DNA

Preliminary experimental data support the idea that therapeutic angiogenesis using short-term gene expression may have the potential to elicit functional improvement in individuals with heart failure (Fig. 2). Previous anatomical studies have established that interstitial fibrosis is a characteristic finding in people with dilated cardiomyopathy⁴⁹. The physiological implication of this finding has been identified by Sabbah *et al.*⁵⁰, who used a canine model of chronic heart failure to show that capillary density is decreased significantly and the diffusion distance of oxygen increased significantly in regions of severe reactive interstitial fibrosis. Light microscopic examination of biopsy specimens retrieved from individuals with dilated cardiomyopathy suggests that such interstitial fibrosis may act as a barrier to the intimate access that cardiac myocytes typically have to adjacent capillaries responsible for myocardial perfusion.

Myocardial perfusion defects observed on nuclear perfusion scans — even in the absence of atherosclerotic narrowing of the extramural coronary arteries — support the notion that such

encasing foci of fibrosis functionally segregate the viable cardiac myocytes from the coronary microcirculation. These observations have been reinforced by van den Heuvel *et al.*⁵¹, who showed that there is impaired myocardial blood flow reserve in individuals with idiopathic dilated cardiomyopathy. The impairment was related to the severity of heart failure and to systolic left ventricular wall stress. In addition, in most of these individuals (76%), mismatch (perfusion/metabolism) was observed throughout a substantial part of the myocardium. These *in vivo* findings were interpreted as evidence of a "stress-induced imbalance between oxygen demand (oxygen utilization) and supply, leading to regional metabolic changes compatible with hibernation or chronic ischaemia"⁵¹.

Finally, Carmeliet *et al.*⁵² have generated mice that express only the 120 isoform of VEGF and have depressed left ventricular systolic pressure and contractility, enlarged left ventricle dimensions, impaired fractional shortening, and reduced aortic outflow velocity. Capillary density in the VEGF^{120/120} mice is reduced fourfold and results in larger intercapillary distances, suggesting that oxygen delivery to individual VEGF^{120/120} cardiomyocytes is reduced. Evidence that the defects in myocardial angiogenesis in VEGF^{120/120} mice result in myocardial ischaemia include a reduced regional blood flow through the myocardium, electrocardiogram alterations and hypoxic zones. The angiogenic and ischaemic defects are most pronounced in the subendomyocardium, consistent with previous observations that angiogenic sprouting occurs in an epi- to endocardial gradient⁵³. Moreover, in contrast to the normally perfused heart, inhibition of perfusion experienced during systole in collateral-dependent tissue is not compensated by a reverse gradient of flow in diastole.

These findings have been confirmed in an elegant transgenic model developed by Gerber and colleagues⁵⁴ in which deletion of the VEGF-1 gene only in cardiac myocytes was accomplished by mating VEGF-loxP mice with myosin light chain 2v-Cre mice that direct *cre* expression specifically to ventricular cardiomyocytes. The hearts of these animals are characterized by impaired left ventricular contractility associated with a highly statistically significant reduction in coronary microvessels.

These results of gene targeting in mice imply that the coronary microvasculature is responsible for both myocardial ischaemia and

left ventricle dysfunction. Angiographic observations made in the individuals who received phVEGF₁₆₅ gene transfer have likewise suggested that the increased vascularity responsible for the improvements in myocardial perfusion and anginal symptoms occurs in the microvascular coronary circulation. No evidence for new collateral vessels has been documented. Studies of DNA labelling in pig, and canine models of myocardial ischaemia, have established that improvement in collateral-dependent flow typically results from the proliferation of vessels of less than 200 µm in diameter⁵⁵. This observation has been confirmed by the *in vivo* imaging studies of Takeshita *et al.*⁵⁶. However, current imaging techniques are neither suitable to resolve the extent to which deficient microvasculature contributes to the development of a cardiomyopathic ventricle, nor adequate to assess neovascularization in response to therapeutic angiogenesis. It thus remains tenable that, in certain cases of heart failure, the angiographically occult intramural coronary vasculature constitutes the locus of disease pathogenesis as well as the therapeutic target.

Indirect gene transfer

Although myocardial gene transfer in human subjects has been achieved exclusively by directly administering the transgene as plasmid DNA or viral vector, animal studies have established the feasibility, and in some cases the potential advantages, of indirect gene transfer through the administration of cells that have been transfected *ex vivo* before *in vivo* delivery (Fig. 2). Asahara *et al.* have shown that *ex vivo* gene transfer of the murine 164 isoform of VEGF (VEGF₁₆₄) augments the proliferative activity of endothelial progenitor cells (EPCs)⁵⁷ and enhances the adhesion and incorporation of EPCs into quiescent as well as activated endothelial cell monolayers⁵⁸. Previous studies^{57,59–63} have established that bone marrow-derived EPCs present in the systemic circulation both home to and incorporate into sites of neovascularization, and may be useful in therapeutic strategies of 'supply-side angiogenesis'. EPCs gathered from the peripheral circulation, for example, have been expanded *ex vivo* and then administered to animals with limb⁶⁴ or myocardial⁶⁵ ischaemia to enhance neovascularization successfully.

Physiological evidence of neovascular function in these pre-clinical animal models includes a high rate of limb salvage and improvement in myocardial function. EPC transplantation thus constitutes a new therapeutic strategy that might provide a robust source of viable endothelial cells to supplement the contribution of endothelial cells resident in the adult vasculature that migrate, proliferate and remodel in response to angiogenic cues, according to the classic model of angiogenesis developed by Folkman⁶⁶.

Alternatively, indirect gene transfer has been proposed as an adjunct to autologous skeletal myoblast transplantation for ischaemic cardiomyopathy^{67,68}. This approach involves administering more than 8×10^8 cells (estimated to include roughly 65% myoblasts) expanded and grown *ex vivo* from thigh muscle biopsies. Studies performed in animal models have documented that grafting such cells into ischaemic myocardium may improve systolic and diastolic myocardial performance *in vivo*⁶⁹. Although the transplanted myoblasts are likely to secrete a variety of endogenous cytokines, including angiogenic growth factors, *ex vivo* transduction with VEGF has been used to ensure constitutive gene expression and facilitate engraftment⁷⁰. Pre-implantation, *ex vivo* overexpression of connexin-43 has also been discussed as a means of modulating skeletal muscle phenotype to enhance cell–cell interaction with host cardiomyocytes⁶⁷.

Perspectives

Over the past two years, clinical trials of gene therapy have come under intense scrutiny by different federal regulatory agencies, and thus clinical data regarding the efficacy of this therapeutic strategy remain limited. The ability to extend considerably the half-life of the encoded agent, as well as to optimize local concentrations while minimizing systemic exposure, are inherent advantages of gene transfer.

Evidence of efficacy in the investigations carried out so far suggests that these conceptual features may translate into clinical advantages for gene transfer over other conventional alternatives, including the use of recombinant proteins.

It is hoped that the clinical evidence showing that myocardial gene transfer is safe and well tolerated may now permit larger scale, appropriately randomized trials to take place. In this regard, it is noteworthy that many of the potential toxicities highlighted by experiments involving angiogenic growth factors of genetically engineered mice have not been borne out in clinical trials. Specifically, clinical trials do not suggest that the use of angiogenic cytokines accelerates atherosclerosis; indeed, the available data suggest that some angiogenic cytokines, such as VEGF, may have a use in preventing restenosis. This strategy is parenthetically being tested clinically (using gene transfer) for individuals with lower extremity vascular disease, and will soon begin for those with coronary artery disease, with both catheter and stent delivery of the transgene.

Finally, the potential to use cell therapy in a complementary fashion to gene transfer has now been established in pre-clinical studies and is undergoing initial trials in humans. This 'supply-side' strategy may permit the replacement of senescent endothelial cells and/or cardiac myocytes with robust precursors that can integrate into host tissues, and at the same time deliver with each cell a 'payload' of gene products that can restore physiological function. It is likely that this approach will emerge as an effective option for promoting both lower myocardial revascularization and/or myocyte regeneration. □

1. Takeshita, S. *et al.* Therapeutic angiogenesis: a single intra-arterial bolus of vascular endothelial growth factor augments revascularization in a rabbit ischemic hindlimb model. *J. Clin. Invest.* **93**, 662–670 (1994).
2. Isner, J. M. *et al.* Arterial gene therapy for therapeutic angiogenesis in patients with peripheral artery disease. *Circulation* **91**, 2687–2692 (1995).
3. Isner, J. M. *et al.* Clinical evidence of angiogenesis following arterial gene transfer of phVEGF₁₆₅. *Lancet* **348**, 370–374 (1996).
4. Baumgartner, I. *et al.* Constitutive expression of phVEGF₁₆₅ following intramuscular gene transfer promotes collateral vessel development in patients with critical limb ischaemia. *Circulation* **97**, 1114–1123 (1998).
5. Losordo, D. W. *et al.* Gene therapy for myocardial angiogenesis: initial clinical results with direct myocardial injection of phVEGF₁₆₅ as sole therapy for myocardial ischaemia. *Circulation* **98**, 2800–2804 (1998).
6. Losordo, D. W. *et al.* Use of the rabbit ear artery to serially assess foreign protein secretion after site specific arterial gene transfer *in vivo*: evidence that anatomic identification of successful gene transfer may underestimate the potential magnitude of transgene expression. *Circulation* **89**, 785–792 (1994).
7. Rosengart, T. K. *et al.* Angiogenesis gene therapy: phase I assessment of direct intramyocardial administration of an adenovirus vector expression VEGF121 cDNA to individuals with clinically significant severe coronary artery disease. *Circulation* **100**, 468–474 (1999).
8. Grimes, C. Adenovirus FGF angiogenic gene therapy (AGENT) trial. Presented at Late-Breaking Clinical Trials session of 50th annual American College of Cardiology, Orlando, FL, 19 March (2001).
9. Wolff, J. A. *et al.* Direct gene transfer into mouse muscle *in vivo*. *Science* **247**, 1465–1468 (1990).
10. Takeshita, S., Isshiki, T. & Sato, T. Increased expression of direct gene transfer into skeletal muscles observed after acute ischemic injury in rats. *Lab. Invest.* **74**, 1061–1065 (1996).
11. Gal, D. *et al.* Direct myocardial transfection in two animal models: evaluation of parameters affecting gene expression and percutaneous gene delivery. *Lab. Invest.* **68**, 18–25 (1993).
12. Rauh, G., Pieczek, A., Irwin, W., Schainfield, R. & Isner, J. M. *In vivo* analysis of intramuscular gene transfer in human subjects studied by on-line ultrasound imaging. *Hum. Gene Ther.* (in the press).
13. Tio, R. A. *et al.* Intramyocardial gene therapy with naked DNA encoding vascular endothelial growth factor improves collateral flow to ischemic myocardium. *Hum. Gene Ther.* **10**, 2953–2960 (1999).
14. Esakof, D. D. *et al.* Intraoperative multiphase transesophageal echocardiography for guiding direct myocardial gene transfer of vascular endothelial growth factor in patients with refractory angina pectoris. *Hum. Gene Ther.* **10**, 2315–2323 (1999).
15. Hartikka, J. *et al.* An improved plasmid DNA expression vector for direct injection into skeletal muscle. *Hum. Gene Ther.* **7**, 1205–1217 (1996).
16. Schratzberger, P. *et al.* Ultrasound enhances therapeutic gene expression in ischemic pig myocardium. *J. Am. College Cardiol.* **37**, 266A (2001).
17. Taniyama, Y. *et al.* Local delivery of naked plasmid DNA of P53 gene into rat balloon injured carotid artery using low voltage ultrasound with contrast microbubbles (option): development of novel non-viral transfection system into blood vessels. *Circulation* **102**, II-164 (2000).
18. Risau, W. Differentiation of endothelium. *FASEB J.* **9**, 926–933 (1995).
19. Allen, K. B. *et al.* Comparison of transmyocardial revascularization with medical therapy in patients with refractory angina. *N. Engl. J. Med.* **341**, 1029–1036 (1999).
20. Symes, J. F. *et al.* Gene therapy with vascular endothelial growth factor for inoperable coronary artery disease: preliminary clinical results. *Ann. Thor. Surg.* **68**, 830–837 (1999).
21. Rosengart, T. K. *et al.* Video assisted epicardial delivery of angiogenic gene therapy to the human myocardium utilizing an adenovirus vector encoding for VEGF₁₂₁. *Circulation* **100**, I-770 (1999).
22. Vincent, K. A. *et al.* Angiogenesis is induced in a rabbit model of hindlimb ischaemia by naked DNA encoding a HIF-1α/VP16 hybrid transcription factor. *Circulation* **102**, 2255–2261 (2000).
23. Iwaguro, H. *et al.* Angiogenesis induced by adenovirus-mediated gene transfer of a hypoxia-inducible factor-1α/VP16 hybrid in rabbit hindlimb ischaemia. *Circulation* **100**, I-47 (1999).

24. Vale, P. R. *et al.* Randomized, single-blind, placebo-controlled pilot study of catheter-based myocardial gene transfer for therapeutic angiogenesis using left ventricular electromechanical mapping in patients with chronic myocardial ischaemia. *Circulation* **103**, 2138–2143 (2001).
25. Shah, A. S. *et al.* *In vivo* ventricular gene delivery of a β -adrenergic receptor kinase inhibitor to the failing heart reverses cardiac dysfunction. *Circulation* **103**, 1311–1316 (2001).
26. Seidman, J. G. & Seidman, C. The genetic basis for cardiomyopathy: from mutation identification to mechanistic paradigms. *Cell* **104**, 557–567 (2001).
27. Schmidt, U. *et al.* Contribution of abnormal sarcoplasmic reticulum ATPase activity to systolic and diastolic dysfunction in human heart failure. *J. Mol. Cell Cardiol.* **30**, 1929–1937 (1998).
28. Miyamoto, M. I. *et al.* Adenoviral gene transfer of SERCA2a improves left ventricular function in aortic-banded rats in transition to heart failure. *Proc. Natl Acad. Sci. USA* **97**, 793–798 (2000).
29. Hajjar, R. J., del Monte, F., Matsui, T. & Rosenzweig, A. Prospects for gene therapy for heart failure. *Circ. Res.* **86**, 616–621 (2000).
30. He, H. *et al.* Overexpression of the rat sarcoplasmic reticulum Ca^{2+} ATPase gene in the heart of transgenic mice accelerates calcium transients and cardiac relaxation. *J. Clin. Invest.* **100**, 380–389 (1997).
31. Periasamy, M. Adenoviral-mediated SERCA gene transfer into cardiac myocytes. How much is too much? *Circulation* **88**, 373–375 (2001).
32. He, H. *et al.* Effects of mutant and antisense RNA of phospholamban on SR Ca^{2+} -ATPase activity and cardiac myocyte contractility. *Circulation* **100**, 974–980 (1999).
33. Lefkowitz, R. J., Rockman, H. A. & Koch, W. J. Catecholamines, cardiac “ β ” adrenergic receptors, and heart failure. *Circulation* **101**, 1634–1637 (2000).
34. Maurice, J. P. *et al.* Enhancement of cardiac function after adenoviral-mediated *in vivo* intracoronary β 2-adrenergic receptor gene delivery. *J. Clin. Invest.* **104**, 21–29 (1999).
35. Weig, H.-J. *et al.* Enhanced cardiac contractility after gene transfer of V2 vasopressin receptors *in vivo* by ultrasound-guided injection of transcoronary delivery. *Circulation* **101**, 1578–1585 (2000).
36. White, D. C., Hata, J. A. & Shah, A. S. Preservation of β -adrenergic receptor signaling delays the development of heart failure after myocardial infarction. *Proc. Natl Acad. Sci. USA* **97**, 5428–5433 (2000).
37. Lai, N. C. *et al.* Intracoronary delivery of adenovirus encoding adenylyl cyclase VI increases left ventricular function and cAMP-generating capacity. *Circulation* **102**, 2396–2401 (2000).
38. Haunsettinger, A. & Izumo, S. Toward antiapoptosis as a new treatment modality. *Circ. Res.* **86**, 371–376 (2000).
39. Hirota, H. *et al.* Loss of gp130 cardiac muscle cell survival pathway is a critical event in the onset of heart failure during biomechanical stress. *Cell* **97**, 189–198 (1999).
40. Krishenbaum, L. A. & de Moissac, D. The bcl-2 gene product prevents programmed cell death of ventricular myocytes. *Circulation* **96**, 1580–1585 (1997).
41. Matsui, T. *et al.* Adenoviral gene transfer of activated PI 3-kinase and Akt inhibits apoptosis of hypoxic cardiomyocytes *in vitro*. *Circulation* **275**, 661–665 (1999).
42. Fujio, Y., Nguyen, T., Wencker, D., Kitsis, R. N. & Walsh, K. Akt promotes survival of cardiomyocytes *in vitro* and protects against ischaemia-reperfusion injury in mouse heart. *Circulation* **101**, 660–667 (2000).
43. Mack, C. A. *et al.* Biologic bypass with the use of adenovirus-mediated gene transfer of the complementary deoxyribonucleic acid for vascular endothelial growth factor 121 improves myocardial perfusion and function in the ischemic porcine heart. *J. Thor. Cardiovasc. Surg.* **115**, 168–176 (1998).
44. Giordano, F. J. *et al.* Intracoronary gene transfer of fibroblast growth factor-5 increases blood flow and contractile function in an ischemic region of the heart. *Nature Med.* **2**, 534–539 (1996).
45. O'Donnell, J. M. *et al.* Tight control of exogenous SERCA expression is required to obtain acceleration of calcium transients with minimal cytotoxic effects in cardiac myocytes. *Circ. Res.* **88**, 415–421 (2001).
46. Svensson, E. C. *et al.* Efficient and stable transduction of cardiomyocytes after intramyocardial injection or intracoronary perfusion with recombinant adeno-associated virus vectors. *Circulation* **99**, 201–205 (1999).
47. Su, H., Lu, R. & Kan, Y. W. Adeno-associated viral vector-mediated vascular endothelial growth factor gene transfer induces neovascular formation in ischemic heart. *Proc. Natl Acad. Sci. USA* **97**, 13801–13806 (2000).
48. Blau, H. M. & Banfi, A. The well-tempered vessel. *Nature Med.* **7**, 532–534 (2001).
49. Weber, K. T. *et al.* Fibrillar collagen and remodeling of dilated canine left ventricle. *Circulation* **82**, 1387–1401 (1990).
50. Sabbah, H. N., Sharov, V. G., Lesch, M. & Godstein, S. Progression of heart failure: a role for interstitial fibrosis. *Mol. Cell Biochem.* **147**, 29–34 (1995).
51. van den Heuvel, A. F. M. *et al.* Regional myocardial blood flow reserve impairment and metabolic changes suggesting myocardial ischaemia in patients with idiopathic dilated cardiomyopathy. *J. Am. Coll. Cardiol.* **35**, 19–28 (2000).
52. Carmeliet, P. *et al.* Impaired myocardial angiogenesis and ischemic cardiomyopathy in mice lacking the vascular endothelial growth factor isoforms VEGF₁₈₈. *Nature Med.* **5**, 495–502 (1999).
53. Tomanek, R. J. Formation of the coronary vasculature: a brief review. *Cardiovasc. Res.* **31**, E46–E51 (1996).
54. Gerber, H. P. *et al.* Knock out of VEGF in ventricular cardiomyocytes leads to increased embryonic lethality and severe cardiac defects in surviving mice. Abstr. Poster Presentation 120, 2000 Keystone Conference on Angiogenesis, 3 March 2000, Salt Lake City, UT (2000).
55. White, F. C., Carroll, S. M., Magnet, A. & Bloor, C. M. Coronary collateral development in swine after coronary artery occlusion. *Circ. Res.* **71**, 1490–1500 (1992).
56. Takeshita, S. *et al.* Use of synchrotron radiation microangiography to assess development of small collateral arteries in a rat model of hindlimb ischaemia. *Circulation* **95**, 805–808 (1997).
57. Asahara, T. *et al.* Isolation of putative progenitor endothelial cells for angiogenesis. *Science* **275**, 964–967 (1997).
58. Asahara, T. *et al.* Gene therapy of endothelial progenitor cell for vascular development in severe ischemic disease. *Circulation* **100**, 1–481 (1999).
59. Asahara, T. *et al.* Bone marrow origin of endothelial progenitor cells responsible for postnatal vasculogenesis in physiological and pathological neovascularization. *Circ. Res.* **85**, 221–228 (1999).
60. Takahashi, T. *et al.* Ischaemia- and cytokine-induced mobilization of bone marrow-derived endothelial progenitor cells for neovascularization. *Nature Med.* **5**, 434–438 (1999).
61. Kalka, C. *et al.* Mobilization of endothelial progenitor cells following gene therapy with VEGF₁₆₅ in patients with inoperable coronary disease. *Ann. Thor. Surg.* **70**, 829–834 (2000).
62. Shi, Q. *et al.* Evidence for circulating bone marrow-derived endothelial cells. *Blood* **92**, 362–367 (1998).
63. Crosby, J. R. *et al.* Endothelial cells of hematopoietic origin make a significant contribution to adult blood vessel formation. *Circ. Res.* **87**, 728–730 (2000).
64. Kalka, C. *et al.* Transplantation of *ex vivo* expanded endothelial progenitor cells for therapeutic neovascularization. *Proc. Natl Acad. Sci. USA* **97**, 3422–3427 (2000).
65. Kawamoto, A. *et al.* Therapeutic potential of *ex vivo* expanded endothelial progenitor cells for myocardial ischaemia. *Circulation* **103**, 634–637 (2001).
66. Folkman, J. Tumor angiogenesis: therapeutic implications. *N. Engl. J. Med.* **285**, 1182–1186 (1971).
67. Menache, P. Autologous skeletal myoblast transplantation for ischemic cardiomyopathy: first clinical case. *Cardiac Vasc. Regen.* **1**, 155–156 (2000).
68. Ellis, M. J., Russell, S. D. & Taylor, D. A. Translating cell transfer for cardiovascular disease to the bedside: a preclinical review and discussion of potential early trials. *Cardiac Vasc. Regen.* **1**, 197–204 (2000).
69. Taylor, D. A. *et al.* Regenerating functional myocardium: improved performance after skeletal transplantation. *Nature Med.* **4**, 929–933 (1998).
70. Cherwek, D., Hopkins, M. B., Hutcheson, K. A., Urbaniak, J. R. & Taylor, D. A. Relieving exercise intolerance secondary to heart failure: myoblast-mediated angiogenesis via VEGF delivery to ischemic skeletal muscle. *Circulation* **100**, I-657 (1999).

Myocyte renewal and ventricular remodelling

Piero Anversa & Bernardo Nadal-Ginard

Cardiovascular Research Institute, Department of Medicine, New York Medical College, Valhalla, New York 10595, USA
(e-mail: piero_anversa@nymc.edu)

Remaining young at heart is a desirable but elusive goal. Unbeknown to us, however, myocyte regeneration may accomplish just that. Continuous cell renewal in the adult myocardium was thought to be impossible, but multipotent cardiac stem cells may be able to renew the myocardium and, under certain circumstances, can be coaxed to repair the broken heart after infarction.

Cardiac myocytes are thought to be terminally differentiated cells and have been often compared to neurons for their inability to regenerate and replace damaged myocardium. Even though evidence now exists for adult neurogenesis and neural stem cells¹, the concept of myocyte regeneration has not been embraced by the medical community and remains highly disputed². Hypertrophy has been assumed to be the only form of myocyte growth in the heart, and over the years information has been gathered on the many signalling pathways implicated in myocyte hypertrophy³. Conversely, the mechanisms of myocyte regeneration have been mostly neglected.

Difficulties in interpreting cellular labelling experiments — owing to the complexity of identifying myocytes that are duplicating DNA by conventional light microscopy — have limited analysis of myocyte regeneration. But recent results in humans and animals have provided evidence that myocyte replication does occur under physiological and pathological conditions of the heart^{4–6}. The use of Ki67 and 5-bromodeoxyuridine (BrdU) as markers of cell proliferation, together with the use of contractile protein antibodies for recognizing myocytes, has allowed us to identify multiplying myocytes by high-resolution confocal microscopy^{4–6}. In spite of these observations, however, the scepticism about myocyte division remains so strong that the upregulation of cyclins, cyclin-dependent kinases (CDKs) and telomerase activity in the heart have been viewed as biochemical events of cellular hypertrophy^{7,8} rather than indices of cell proliferation.

Myocyte replication and the infarcted heart

Evidence suggesting that some myocytes divide comes from several studies in animal models that show that myocytes express early and late growth-related genes immediately after infarction. Quantities of cyclin E, A and B are increased and their associated kinase activities are elevated significantly⁴. In addition, high levels of DNA replication, karyokinesis and cytokinesis have been identified^{4–6}. These aspects of ventricular remodelling have been shown to occur after infarction, and in other models of heart failure in which mitotic indices have been measured in myocardial sections and dissociated myocytes^{5,9}. The concept of multiple myocyte divisions in the mammalian heart has been strengthened by the recognition of telomeric shortening and the decrease of

telomerase activity^{10,11} in the decompensated ageing rat heart. However, acute heart failure in dogs is characterized by cell regeneration with preservation of telomeric length, owing to a marked increase in the activity of telomerase¹².

Studies of the postinfarcted human heart shortly after coronary occlusion and late during the terminal stages of the ischaemic myopathy have characterized the effects of time on the extent of myocyte replication^{5,6}. Notably, myocytes in mitosis are present in control hearts, which suggests that myocyte regeneration contributes to the homeostasis of the nondiseased heart. Cell growth is markedly enhanced acutely after infarction and more in the border zone than in the remote tissue. The number of dividing myocytes (Fig. 1a–c) is 3–4-fold higher at 1 week after infarction⁶ than in end-stage cardiac failure, years after the primary event⁵. Because myocytes in the infarcted area die in a few hours and ischaemic damage destroys the vascular and nonvascular components of the interstitium, formation of new myocardium in the infarcted region through myocyte growth alone would seem to be impossible. Cell proliferation occurs exclusively in the border zone and in distant tissue where the blood supply is largely maintained¹³.

Quantitative results consistent with myocyte proliferation have been frequently based on the assumption that myocytes are mononucleated cells^{2,4}. Thus, an increase in the total number of myocyte nuclei would correspond to an identical increase in the number of cells in the ventricle. But different proportions of mononucleated and binucleated myocytes may be present, and this possibility complicates the distinction between karyokinesis and cytokinesis. Karyokinesis without cytokinesis results in an increase in the number of nuclei per cell, but the actual number of cells remains constant. Conversely, cytokinesis produces an increase in the number of cells, whereas the number of nuclei per cell does not change. Importantly, the human heart is composed of nearly 80% mononucleated and 20% binucleated ventricular myocytes, and this ratio is not altered by sex, ageing, cardiac hypertrophy or ischaemic cardiomyopathy¹⁴.

Together, these data support the notion that a subpopulation of adult myocytes re-enter the cell cycle and proliferate. Myocyte regeneration and cellular hypertrophy constitute the growth reserve of the heart and can expand significantly the functioning myocardium after infarction (Fig. 1d). Thus, the concept that myocardial infarction

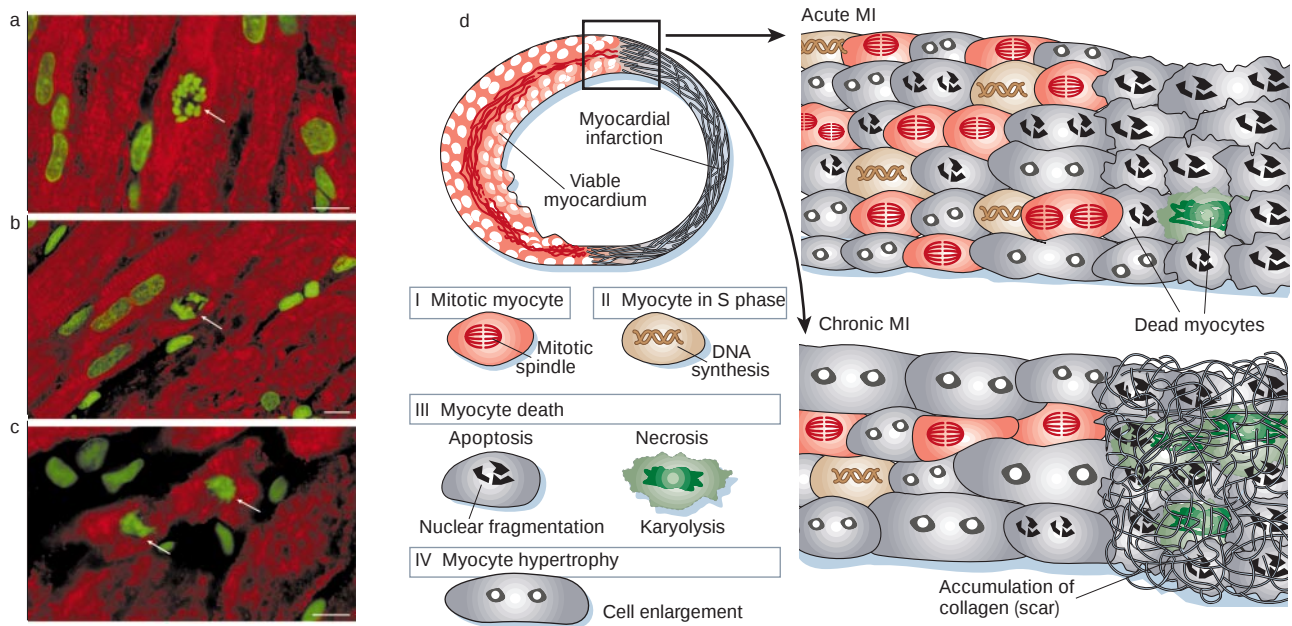


Figure 1 Myocyte division. **a–c**, Stages of mitosis in human myocytes. Metaphase chromosomes (**a**), karyokinesis (**b**) and cytokinesis (**c**) are shown by the green fluorescence of propidium iodide (PI). Red fluorescence reflects cardiac myosin

antibody staining of myocyte cytoplasm. Arrows indicate mitosis. Scale bars, 10 μ m. **d**, Changes in myocyte growth in acute and chronic infarcts. MI, myocardial infarct.

represents a demonstration of the terminally differentiated state of myocytes should be reconsidered.

Origin of replicating myocytes

The identification of cycling cells in the myocardium as myocytes is based on their expression of myocyte-specific markers. To facilitate their recognition, interstitial cells are not labelled and appear as scattered nuclei (Fig. 1a–c). Components of the contractile apparatus have been the most commonly used indicators of the origin of these dividing cells^{4–6}. Myocytes in mitosis may be rich in organized myofibrils that occupy a predominant portion of the cytoplasm⁶. These characteristics indicate that there are mature myocytes in the adult human heart that have the ability to re-enter the cell cycle and divide. However, the extent of the replicative potential of myocytes is unknown.

There are two possible origins of cycling myocytes. First, these cells might be part of a small pool of replicating myoblasts that divide asymmetrically and continuously generate new myocytes for normal homeostasis or in response to pathological stimuli. But these cells do not grow in culture, which, together with the absence of evidence for clonal expansion *in vivo* in BrdU-labelling experiments in rats¹⁵,

makes the existence of such a population unlikely. Second, the dividing myocytes might be amplifying cells derived from stem cells that expand and produce a differentiated progeny under proper stimulation. This progeny could represent the cycling myocytes that after a few cell divisions withdraw from the cell cycle and become terminally differentiated, reaching growth arrest⁶.

An important issue is whether myocyte precursors derive from resident cardiac stem cells (CSCs) or from circulating stem cells that have homed to the heart. Undifferentiated cells expressing stem-cell-related antigens have been identified in the adult myocardium. Our unpublished data show that primitive cells can be detected by three surface markers (Fig. 2): *c-kit*, which is the receptor^{16,17} for stem cell factor; MDR1, which is a P-glycoprotein capable of extruding dyes, toxic substances and drugs^{18,19}; and Sca-1, which is involved in cell signalling and cell adhesion²⁰. None of these markers is specific for stem cells (Fig. 3), as each one is found in haematopoietic stem cells and other cell types^{21–24}. Whether these cells are CSCs remains to be ascertained. But *Lin[−]c-kit^{POS}* bone marrow cells can reconstitute mouse myocardium *in vivo*^{25,26}.

An issue of biological and clinical relevance is whether newly formed myocytes are derived from CSCs that accumulated in the

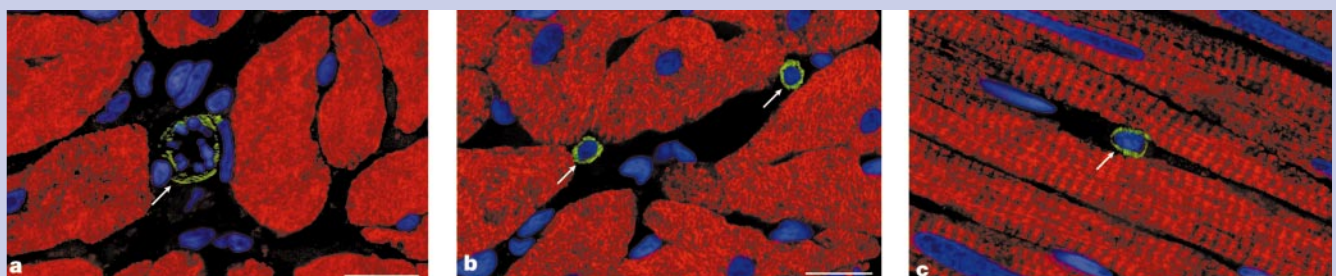


Figure 2 Primitive cells and myocyte ageing. Cells are positive (green) for *c-kit* (**a**), MDR1 (**b**) and Sca-1 (**c**) in rat ventricular myocardium. Cardiac myosin is stained

red, PI is blue. The *c-kit*-labelled cell is in mitosis (**a**). Scale bars, 10 μ m.

heart early in development or are the progeny of haematopoietic stem cells that, later in life, home to the myocardium from the systemic circulation. In favour of the first possibility is the fact that *c-kit*^{POS} cells migrate during fetal growth and form colonies in several organs, including the heart^{27,28}. Chemotaxis of haematopoietic stem cells is modulated by stem cell factor, which promotes their migration to specific sites²⁹, suggesting that stem cells may have been stored in the heart as remnants from the cardiac primordia. These primitive cells may have undergone symmetric and asymmetric division³⁰, expanding the pool size of undifferentiated cells in the maturing heart. The recent studies by us and our co-workers^{25,26} and others^{31–33} showing that bone marrow cells (BMCs) can regenerate myocardium after infarction establish that blood-borne cells can acquire properties of CSCs and differentiate.

Further support for the role of adult cycling myocytes in ventricular remodelling has come from the recognition that 20% of these cells in dogs express telomerase¹². Many telomerase-competent myocytes are cycling, as indicated by the presence of Ki67 protein in their nuclei; however, the number of divisions may be limited. Telomeres constitute the physical ends of chromosomes, and telomerase can keep the length of telomeres intact after each cell cycle^{34,35}. This prevents premature senescence and sustains cell multiplication³⁶. Progenitor cells and rapidly replicating, amplifying cells have high levels of telomerase activity³⁷. Telomerase has been recognized not only in myocytes but also in neurons from organs with low turnover rates^{12,38}.

Myocyte ageing, volume and growth

It remains a general belief that the number of myocytes in the heart is defined at birth and these cells persist throughout life. There are men and women 100 years old and older, and this fact would imply that all of their myocytes have lived 100 years or more — in other words, the age of individuals and the age of their myocytes should coincide. But myocytes do not live indefinitely — they have a limited lifespan in humans and rodents^{39,40}. Cell loss and myocyte proliferation are part and parcel of normal homeostasis, and an increase in these parameters is typical of cardiac ageing⁴¹. The old heart is characterized by a reduction in cell number and hypertrophy of the remaining myocytes⁴¹, and this phenotype has been used to argue against the formation of new myocytes. But without cell regeneration the rates of cell death^{4,41} would imply that all myocytes would die during the first few months of a rodent's lifespan⁴⁰. For example, the left ventricle of a young rat contains 13×10^6 myocytes, and at any point in time

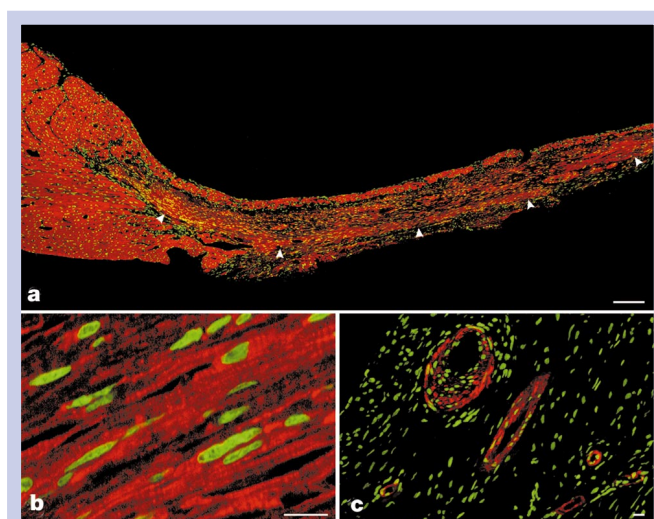


Figure 4 Myocardial repair. **a**, Band of regenerating myocardium after infarction (arrowheads). **b, c**, Newly formed small myocytes (**b**) and coronary arterioles (**c**) in the developing band. Cardiac myosin (**a, b**) and α -smooth muscle actin (**c**) are stained red, PI is yellow-green. Scale bars, 100 μ m (**a**); 10 μ m (**b, c**).

200 and 93,000 myocytes are dying by apoptosis and necrosis, respectively. Because apoptosis is completed in nearly 4 h and necrosis in roughly 24 h, 94,200 myocytes are lost in one day. Thus, 2.83×10^6 cells would die in 1 month, and the total 13×10^6 ventricular myocytes would disappear in 5 months.

Throughout life, a mixture of young and old cells is present in the rat myocardium. Although most myocytes seem to be terminally differentiated, there is a fraction of younger myocytes (15–20%) that retains the capacity to replicate^{10,11}. The proportion of these two populations changes with age, and there is no point in life at which all myocytes are comparable in terms of age, size, shape and molecular properties. Whether a myocyte responds to pathological loads by hypertrophy or replication is influenced by cell volume, which in turn reflects its age. Large cells are old, do not react to growth stimuli, and are more prone to activate cell death. Small cells are younger, can re-enter the cell cycle or hypertrophy, and are less susceptible to death.

There is a good correlation between myocyte age and expression of the CDK inhibitor p16^{INK4a} (p16), which has proved to be a marker of cellular ageing⁴². p16 is detectable in 10% of myocytes at birth, and in more than 80% of myocytes in senescent rats¹⁰. Telomeric length follows a similar pattern (longer in younger, smaller cells and shorter in older, larger ones), reflecting a decrease of telomerase with age^{10,11}. The age-dependent increase in myocyte death, coupled with a reduction in coronary vasculature, may explain why myocardial infarction is associated with increased mortality in the elderly⁴³.

Cellular therapy for myocardial infarction

New therapies of myocardial infarction include the implantation of skeletal myoblasts and bone-marrow-derived cardiomyocytes^{44,45}; however, these strategies have failed to reconstitute healthy myocardium⁴⁶. The growth potential of adult BMCs offers a promising new tool. These cells home to the infarcted region by local injection²⁵, by mobilization by cytokines²⁶ or by spontaneous translocation after injury³². Homed BMCs proliferate and differentiate into myocytes, smooth muscle cells and endothelial cells, resulting in the partial regeneration of the destroyed myocardium^{25,26}. In addition, the growth response mediated by BMCs interferes with ventricular scarring and decompensation^{25,26}. Small myocytes and vascular structures develop in the infarct and mostly replace the dead zone (Fig. 4). For the first time, BMCs have been shown to generate myocardium *in vivo*, thus reducing infarct size^{25,26}.

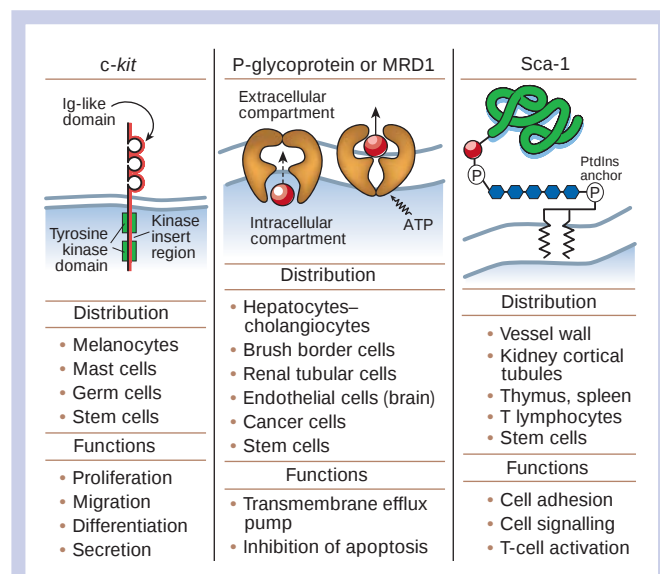


Figure 3 Structure, distribution and function of surface markers of stem cells.

Contraction reappears in the area of injury, diastolic stress is decreased, ventricular haemodynamics is improved and mortality is reduced²⁶. The new myocytes are more reminiscent of fetal than adult cells, however, and the chronic evolution of the reconstituted heart remains to be determined.

But should BMCs be considered to be the cells of choice for cardiac repair or could CSCs be used to replace necrotic myocardium? CSCs might be more effective than BMCs in rebuilding dead tissue. BMCs have to re-programme themselves to give rise to progeny differentiating into cardiac lineages; by contrast, activation and migration of CSCs to the site of injury would avoid this intermediate phase. Moreover, CSCs may be faster than BMCs in reaching functional competence and structural characteristics of mature myocytes and vessels. This approach would require the recognition of growth factors that affect CSCs in a discrete and restrictive manner, and do not stimulate BMCs or stem cells in other organs. The identification of cardiac cells with surface markers of stem cells (Fig. 2) is consistent with this possibility. The high degree of myocyte regeneration in the non-infarcted myocardium in humans⁶ supports this hypothesis. Recruitment and expansion of CSCs would promote a pool of young replicating myocytes and neovascularization.

Future directions

Here we have outlined several new findings that provide an alternative view of myocardial biology that might one day lead to a reconsideration of the long-term goals of the therapy of myocardial infarction and chronic heart failure. This outlook advances the possibility that it might be feasible to develop strategies for the actual regeneration of the infarcted ventricle.

In terms of the controversial proposal that the normal heart undergoes a continuous turnover of myocytes that increases under pathological conditions, we need to understand better the origin and behaviour of cycling myocytes to maximize myocardial growth and cardiac repair. New data suggest that CSCs exist and are the source of tissue renewal. The identification, localization and purification of CSCs, and the characterization of CSC biology, are essential issues that must be resolved. This may lead to the discovery of mediators of CSC migration, proliferation and differentiation that, in turn, might result in the mending of the 'broken heart'.

1. Horner, P. J. & Gage, F. H. Regenerating the damaged central nervous system. *Nature* **407**, 963–970 (2000).
2. Soonpaa, M. H. & Field, L. I. Survey of studies examining mammalian cardiomyocyte DNA synthesis. *Circ. Res.* **83**, 15–26 (1998).
3. Molkentin, J. D. & Dorn, G. W. Cytoplasmic signaling pathways that regulate cardiac hypertrophy. *Annu. Rev. Physiol.* **63**, 391–426 (2001).
4. Anversa, P. & Kajstura, J. Ventricular myocytes are not terminally differentiated in the adult mammalian heart. *Circ. Res.* **83**, 1–14 (1998).
5. Kajstura, J. et al. Myocyte proliferation in end-stage cardiac failure in humans. *Proc. Natl Acad. Sci. USA* **95**, 8801–8805 (1998).
6. Beltrami, A. P. et al. Evidence that human cardiac myocytes divide after myocardial infarction. *N. Engl. J. Med.* **344**, 1750–1757 (2001).
7. Li, J. M., Poolman, R. A. & Brooks, G. Role of G1 phase cyclins and cyclin-dependent kinases during cardiomyocyte hypertrophic growth in rats. *Am. J. Physiol.* **275**, H814–H822 (1998).
8. Oh, H. et al. Telomerase reverse transcriptase promotes cardiac muscle cell proliferation, hypertrophy and survival. *Proc. Natl Acad. Sci. USA* **98**, 10308–10313 (2001).
9. Setoguchi, M. et al. Activation of cyclins and cyclin-dependent kinases, DNA synthesis, and myocyte mitotic division in pacing-induced heart failure in dogs. *Lab. Invest.* **79**, 1545–1558 (1999).
10. Kajstura, J. et al. Telomere shortening is an in vivo marker of myocyte replication and aging. *Am. J. Pathol.* **156**, 813–819 (2000).
11. Leri, A., Malhotra, A., Liew, C.-C., Kajstura, J. & Anversa, P. Telomerase activity in rat cardiac myocytes is age and gender dependent. *J. Mol. Cell. Cardiol.* **32**, 385–390 (2000).
12. Leri, A. et al. Telomerase expression and activity are coupled with myocyte proliferation and preservation of telomeric length in the failing heart. *Proc. Natl Acad. Sci. USA* **98**, 8626–8631 (2001).
13. Hirzel, H. O., Nelson, G. R., Sonnenblick, E. H. & Kirk, E. S. Redistribution of collateral blood flow from necrotic to surviving myocardium following coronary occlusion in the dog. *Circ. Res.* **39**, 214–222 (1976).
14. Olivetti, G. et al. Aging, cardiac hypertrophy and ischemic cardiomyopathy do not affect the proportion of mononucleated and multinucleated myocytes in the human heart. *J. Mol. Cell. Cardiol.* **28**, 1463–1477 (1996).
15. Kajstura, J. et al. Myocyte cellular hyperplasia and myocyte cellular hypertrophy contribute to chronic ventricular remodeling in coronary artery narrowing-induced cardiomyopathy in rats. *Circ. Res.* **74**, 383–400 (1994).
16. Lyman, S. D. & Jacobsen S. E. c-kit ligand and Flt3 ligand: stem/progenitor cell factors with overlapping yet distinct activities. *Blood* **91**, 1101–1134 (1998).
17. Jiang, X. et al. Structure of the active core of human stem cell factor and analysis of binding to its receptor Kit. *EMBO J.* **19**, 3192–3203 (2000).
18. Bunting, K. D., Zhou, S., Lu, T. & Sorrentino B. P. Enforced P-glycoprotein pump function in murine bone marrow cells results in expansion of side population stem cells *in vitro* and repopulating cells *in vivo*. *Blood* **96**, 902–909 (2000).
19. Bakos, E. et al. Characterization of the amino-terminal regions in the human multidrug resistance protein. *J. Cell Sci.* **113**, 4451–4461 (2000).
20. van der Rijn, M., Heimfeld, S., Spangrude, G. J. & Weissman I. L. Mouse hematopoietic stem-cell antigen Sca-1 is a member of the Ly-6 antigen family. *Proc. Natl Acad. Sci. USA* **86**, 4634–4638 (1989).
21. Kissel, H. et al. Point mutation in Kit receptor tyrosine kinase reveals essential roles for Kit signaling in spermatogenesis and oogenesis without affecting other Kit responses. *EMBO J.* **19**, 1312–1326 (2000).
22. Demeule, M., Labelle, M., Regina, A., Berthelet, F. & Beliveau, R. Isolation of endothelial cells from brain, lung and kidney: expression of the multidrug resistance P-glycoprotein isoforms. *Biochem. Biophys. Res. Commun.* **281**, 827–834 (2001).
23. Geick, A., Eichelbaum, M. & Burk, O. Nuclear receptor response elements mediate induction of intestinal MDR1 by rifampin. *J. Biol. Chem.* **276**, 14581–14587 (2001).
24. English, A., Kosoy, R., Pawlinski, R. & Bamezai, A. A monoclonal antibody against the 66-kDa protein expressed in mouse spleen and thymus inhibits Ly-6A. 2-dependent cell–cell adhesion. *J. Immunol.* **165**, 3763–3771 (2000).
25. Orlic, D. et al. Bone marrow cells regenerate infarcted myocardium. *Nature* **410**, 701–705 (2001).
26. Orlic, D. et al. Mobilized bone marrow cells repair the infarcted heart improving function and survival. *Proc. Natl Acad. Sci. USA* **98**, 10344–10349 (2001).
27. Teyssier-Le Discorde, M., Prost, S., Nandrot, E. & Kirszenbaum, M. Spatial and temporal mapping of c-kit and its ligand, stem cell factor expression during human embryonic haemopoiesis. *Br. J. Haematol.* **107**, 247–253 (1999).
28. Kunisada, T. et al. Transgene expression of steel factor in the basal layer of epidermis promotes survival, proliferation, differentiation and migration of melanocyte precursors. *Development* **125**, 2915–2923 (1998).
29. van Dijk, T. B. et al. Stem cell factor induces phosphatidylinositol 3'-kinase-dependent Lyn/Tec/Dok-1 complex formation in hematopoietic cells. *Blood* **96**, 3406–3413 (2000).
30. Morrison, S. J., Shah, N. M. & Anderson, D. J. Regulatory mechanisms in stem cell biology. *Cell* **88**, 287–298 (1997).
31. Tomita, S. et al. Autologous transplantation of bone marrow cells improves damaged heart function. *Circulation* **100**, II-247–II-256 (1999).
32. Kocher, A. A. et al. Neovascularization of ischemic myocardium by human bone-marrow-derived angioblasts prevents cardiomyocyte apoptosis, reduces remodeling and improves cardiac function. *Nature Med.* **7**, 430–436 (2001).
33. Jackson, K. A. et al. Regeneration of ischemic cardiac muscle and vascular endothelium by adult stem cells. *J. Clin. Invest.* **107**, 1395–1402 (2001).
34. Greider, C. W. Telomerase activity, proliferation and cancer. *Proc. Natl Acad. Sci. USA* **95**, 90–92 (1998).
35. Martin-Rivera, L., Herrera, E., Albar, J. P. & Blasco, M. A. Expression of mouse telomerase catalytic subunit in embryos and adult tissues. *Proc. Natl Acad. Sci. USA* **95**, 10471–10476 (1998).
36. Vaziri, H. et al. Evidence for a mitotic clock in human hematopoietic stem cells: loss of telomeric DNA with age. *Proc. Natl Acad. Sci. USA* **91**, 9857–9860 (1994).
37. Liu, K. et al. Constitutive and regulated expression of telomerase reverse transcriptase (hTERT) in human lymphocytes. *Proc. Natl Acad. Sci. USA* **27**, 5147–5152 (1999).
38. Mattson, M. & Klapper, W. Emerging roles for telomerase in neuronal development and apoptosis. *J. Neurosci. Res.* **63**, 1–9 (2001).
39. Olivetti, G., Melissari, M., Capasso, J. M., & Anversa, P. Cardiomyopathy of the aging human heart. *Circ. Res.* **68**, 1560–1568 (1991).
40. Kajstura, J. et al. Necrotic and apoptotic myocyte cell death in the aging heart of Fischer 344 rats. *Am. J. Physiol.* **271**, H1215–H1228 (1996).
41. Anversa, P. & Olivetti, G. in *Handbook of Physiology, Section 2: The Cardiovascular System, Volume I: The Heart* (eds Page, E., Fozzard, H. & Solaro, J.) 75–144 (Oxford Univ. Press, New York, 2001).
42. Hara, E. et al. Regulation of p16^{INK4} expression and implication for cell immortalization and senescence. *Mol. Cell. Biol.* **16**, 859–867 (1996).
43. Maggioni A. P. et al. Age-related increase in mortality among patients with first myocardial infarctions treated with thrombolysis. *N. Engl. J. Med.* **329**, 1442–1448 (1993).
44. Taylor, D. A. et al. Regenerating functional myocardium: improved performance after skeletal myoblast transplantation. *Nature Med.* **4**, 929–933 (1998).
45. Menasché, P. et al. Myoblast transplantation for heart failure. *Lancet* **357**, 279–280 (2001).
46. Murray, C. E., Wiseman, R. W., Schwartz, S. M. & Hauschka, S. D. Skeletal myoblast transplantation for repair of myocardial necrosis. *J. Clin. Invest.* **98**, 2512–2523 (1996).

Acknowledgements

Our work is supported by NIH grants.

The Heart and Drug Therapy



AstraZeneca is proud to sponsor this NATURE Insight on the Heart. The company has a longstanding scientific interest in creating better drugs in this area. The available drugs for heart disease were mostly developed around the middle of the last century; no new antiarrhythmic drugs have been developed in the last 30-50 years. The same is true for the thrombo-embolic area (except for the new anti-platelet agents). AstraZeneca has a long-term interest in cardiovascular research with registered compounds such as propranolol, metoprolol, atenolol, lisinopril, candesartan, felodipin and compounds in late clinical development such as the superstatin Crestor and the oral, direct-acting thrombin inhibitor Exanta. We will continue to have a strong program in this area of heart and drug therapy with special interest in the lipid-lowering, antiarrhythmic, antihypertensive and thrombo-embolic areas.

AstraZeneca's interest in antiarrhythmic agents:

Although cardiac arrhythmias are a major health problem in society, efficient and safe pharmacological treatments are still lacking. In the 1980s, the Cardiac Arrhythmia Suppression Trials (CAST) highlighted the proarrhythmic risk and inefficacy of class I (sodium channel blocking) antiarrhythmic agents in patients at relatively low risk for sudden death. These studies resulted in an increased interest in class III (action potential prolonging) antiarrhythmic agents. However, even though efficacy has been proven for different

class III agents, they, like class I agents, have been shown to be proarrhythmic and the initial optimism for them has waned. Clearly medical need is great – this is probably true for ventricular arrhythmias, and is certainly true for supra-ventricular tachyarrhythmias, including both acute termination and prophylaxis against recurrence of atrial fibrillation or atrial flutter.

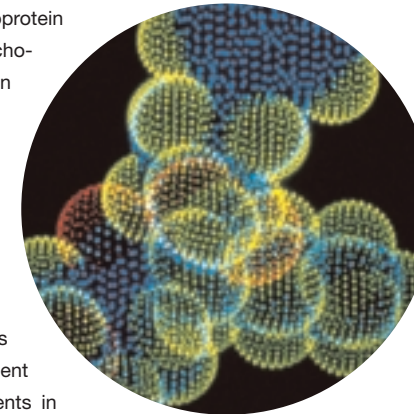
Knowledge about the physiology, structure and molecular biology of cardiac ion channels/ion transporters has grown substantially in recent times. The fruitful combination of electrophysiological techniques with state-of-the-art molecular biology and genetics has not only advanced our understanding of basic cardiac physiology, but also started to provide insights into the pathophysiology of cardiac rhythm disorders. It has become increasingly clear that cardiac arrhythmias can themselves cause drastic changes in electrical, structural and functional properties and that these changes directly promote the maintenance of the arrhythmia.

The rapid progress in this field holds the promise of providing a detailed understanding of the basic determinants of different cardiac arrhythmias in the very near future. It may then be possible to target the development of the arrhythmia substrate (i.e. upstream therapy), rather than modulating its result. Progress in molecular biology and genetics might facilitate prerequisites for more individualised therapy. For example, through genetic screening, it may be possible to identify patients at risk for unwanted side effects or those that will benefit most from a specific therapy. AstraZeneca is at the forefront of using these new approaches to find an effective and, perhaps even more important, safe pharmacological treatment for atrial fibrillation and flutter.

AstraZeneca's interest in vascular disease prevention:

Cardiovascular disease remains the leading cause of death despite the number of drugs with different mechanisms of action used in the clinical setting. Arteriosclerosis in coronary arteries underlying myocardial infarction is a complex trait arising from the interaction of multiple susceptibility genes with a range of environmental factors. Biological remodelling processes in combination with metabolic dysfunction, e.g. insulin resistance and dyslipidemia, cause a set of different plaque architectures in the vessel wall. This ranges from large, fibrotic, stable lesions to lipid-rich unstable plaques, the latter being more prone to rupture and so to causing a thrombotic event. Lipoprotein retention and modification, reverse cholesterol transport, matrix deposition and vascular wall inflammation are all examples of local processes in the atherosclerotic plaque that could be of benefit in treating the disease. Recently it has been elegantly demonstrated by magnetic resonance imaging that lipid lowering treatment with statins decreases the amount of lipids present in the plaque. Technical achievements in non-invasive imaging will enable monitoring of local events in the plaque and make it more attractive to explore antiatherosclerotic therapies in parallel with the more established lipid lowering approaches. AstraZeneca has efficacious lipid-modulating compounds in clinical development and is exploring possible local target mechanisms in the vessel wall. The symptoms of acute plaque rupture and/or erosion with the subsequent exposure of thrombogenic surfaces and formation of thrombi can be

prevented by antithrombotic therapies. AstraZeneca has a commitment to and a leading position in the development of more effective and safer antithrombotic drugs focusing on several mechanisms of the coagulation cascade. The impressive advances in drug discovery and the basic science of cardiac disease will most likely over the next decade translate into significant impact on the clinical therapeutic opportunities available for treatment of cardiovascular diseases.



Peter Thoren
Ingemar Jacobson
Håkan Wennbo
Mikael Dohlsten

*Research Area
Cardiovascular
& Gastrointestinal
AstraZeneca
R&D Mölndal,
S-431 83 Mölndal,
Sweden.*

AstraZeneca

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:

Leonie Welss (4954)

European Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy: Nevin Bayoumi (4978)

Scandinavia: Silje Opstrup (4994)

Spain/ Portugal: Leonie Welss
(4954)

Production Manager:

Billie Franklin

To send films and materials use

London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mails: p.phelan@nature.com

k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakakatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.cowan@naturejp.com

Japan Manager:

Kate Cowan

naturejobs

Moving benchmarks

Levels of support for postdocs in the United States, where biomedical stipends vary between \$25,000 and \$65,000, seem at best to be inscrutable — and at worst simply arbitrary. Whatever the rationale used to set their wages, most postdocs would agree that their salaries are too low, especially when statistics such as median income based on amount of education are included in the equation.

Conventional wisdom suggests that the benchmark for biomedical postdoc stipends is set by the National Institutes of Health (NIH). But this benchmark — which is itself based on another benchmark, the pay of medical residents — is in a constant state of flux, and subject to political and economic pressure.

The political pressure is spearheaded by foundations such as the Howard Hughes Medical Institute, based in Maryland. By offering higher stipends than the government, these organizations hope to drive the benchmark upwards. Meanwhile, industry often offers even higher salaries in a bid to siphon the best and brightest away from the academic world.

The pressure from both sides is coming to bear, as *Naturejobs'* first feature devoted to postdoc and studentship issues reveals (see page 5). The NIH has promised to raise stipends by 10–12% a year over the next few years. This would bring a third-year postdoc's salary in line with that of a first-year medical resident, but would still leave first- and second-year postdocs lagging behind.

What are the ramifications? Industry is likely to continue to outbid government-funded stipends, because the NIH target for experienced postdocs is still below what many companies are willing to pay a promising, if inexperienced, scientist. And foundations will continue their push to move the benchmark higher, especially for less-experienced postdocs.

Paul Smaglik

Naturejobs editor



Contents

POSTDOCS & STUDENTS

Laying the foundations for higher postdoc salaries p5

WWW.NATUREJOBS.COM

Career centre
Information on the scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

Private foundations push for higher postdoc salaries

Michael Teitelbaum:

On average, postdoctoral stipend levels are way too low — they should reflect the levels of education of their recipients.



The levels of most postdoc salaries at US biomedical research institutions are set by the National Institutes of Health. But private philanthropies, which consistently claim such stipends are insufficient, have worked to boost these wages by sponsoring studies on salaries and by offering their own grants with higher levels of support.

“On average we think postdoctoral stipend levels are way too low,” says Michael Teitelbaum, a programme director at the Alfred P. Sloan Foundation, which supports a range of scientific research. “We’d like to see them reflect the sophistication, maturity and the high levels of education of their recipients.”

Foundations have helped to “stimulate a fairer approach to compensating postdocs and I think that’s appropriate”, adds James Gavin, a trustee of the Robert Wood Johnson Foundation, an organization devoted to health and healthcare.

Such stimulation seems to be bearing fruit. Last year, annual salaries for the NIH’s National Research Service Award (NRSA), which funds extramural training grants and fellowships, were \$28,260 for entry-level postdocs, rising to \$44,412 for those with at least seven years’ experience. In March 2001, the NIH promised to include annual cost-of-living increases, and to raise the stipends by 10–12% per year until they reach \$45,000 for new postdocs.

If enacted for this year, these rises will help the NIH to catch up with its own benchmark — house staff salaries at US medical schools, which are now 10–15% higher than the NIH stipends. But any catch-up will not be complete — third year NIH-funded postdocs will achieve parity with their clinical peers, but those with less experience will still lag behind.

A BENEFICIAL ARRANGEMENT

Each foundation has its own approach to making its stipends better reflect a postdoc’s worth. Some offer more benefits, others offer higher pay and some use both incentives. At the Howard Hughes Medical Institute (HHMI) in Maryland, for example, postdocs — known as research associates — are funded through their investigator programme and their salary levels are linked to the NRSA stipends. Minor adjustments are made to that benchmark depending on the cost of

living where the associates work. But because they are considered HHMI employees, the associates receive a healthy benefits package, which is a marked difference from NRSA postdocs, whose benefits are governed by the policy of their home institutions.

Several years ago, when setting postdoc salary levels for its bridging grants in the biomedical sciences, the Burroughs Wellcome Fund, which backs research in the medical sciences, took into account postdoc salaries at organizations such as the NIH, came up with an average, and then exceeded that figure. The salary levels in these grants are reviewed every two to three years. They provide support for up to two years in a postdoc position and three years as a new faculty member, with a total possible award of \$500,000, which also provides a 10% administrative fee to the university to cover health benefits. The maximum salary for the first year is \$38,000, and \$41,000 for the second, with some home institutions supplementing this salary. When grant-holders become faculty members, the grant covers salary and research expenses.

VARIABLE SUPPORT

The Alfred P. Sloan Foundation supports research in theoretical neuroscience through grants to five centres, whose principal investigators in turn support postdocs. Four years ago, Sloan officials told these grantees that they were not comfortable with stipends of less than \$32,000 for any postdoc at any level of experience.

Although stipends vary between foundations, there is more discrepancy among fields and subdisciplines, and bioinformatics is top of the heap. For computational molecular biology, in which Sloan runs a transitional postdoctoral awards programme with the Department of Energy, the stipend is \$50,000, out of which the recipient must buy health benefits.

The Burroughs Wellcome Fund takes the same approach for its ‘interfaces in science’ awards, used to train postdocs from the physical, chemical and computational sciences to apply their skills to biological problems. Stipends are higher to attract people from computer science, who tend to receive higher pay than their peers in biology. But the latter seem at last to be on their way to achieving parity.

Karen Kreeger is a freelance science writer based in Philadelphia.

Web links

NIH research training opportunities

♦ grants.nih.gov/training/nrsa.htm

National Academies’ report

♦ www.nationalacademies.org/postdocs

Howard Hughes Medical Institute

♦ www.hhmi.org

Alfred P. Sloan Foundation

♦ www.sloan.org

Burroughs Wellcome Fund

♦ www.bwffund.org

Robert Wood Johnson Foundation

♦ www.rwjf.org